GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 07:55:20 ; Search time 28.78 Seconds
(without alignments)
2810.584 Million cell updates/sec

Title: US-09-605-783A-113
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	519.5	18.2	599	5	Q9VSV1	Q9vsv1 drosophila
2	488.5	17.1	530	4	Q9UMX9	Q9umx9 homo sapien
3	347.5	12.1	515	10	O65803	O65803 daucus caro
4	347.5	12.1	515	10	Q9FNR6	Q9fnr6 daucus caro
5	346.5	12.1	501	10	Q9SQK6	Q9sqk6 vitis vinif
6	346.5	12.1	515	10	Q9SLN7	Q9sln7 daucus caro
7	346	12.1	612	10	Q9SQK5	Q9sqk5 vitis vinif
8	338	11.8	512	10	Q9S7Z5	Q9s7z5 apium grave
9	337.5	11.8	523	10	O04077	O04077 vicia faba
10	336.5	11.8	501	10	Q9SP63	Q9sp63 vitis vinif
11	333	11.6	512	10	Q9ZTB9	Q9ztb9 apium grave
12	330	11.5	516	10	Q43653	Q43653 solanum tub
13	329.5	11.5	524	10	Q9XHL6	Q9xhl6 pisum sativ
14	328.5	11.5	594	10	O80605	O80605 arabidopsis
15	325.5	11.4	604	10	Q9FVL6	Q9fvl6 lycopersico
16	323.5	11.3	512	10	Q39231	Q39231 arabidopsis
17	323.5	11.3	512	10	O80550	O80550 arabidopsis
18	321.5	11.2	502	10	Q9SP14	Q9sp14 alonsoa mer
19	317.5	11.1	507	10	Q40583	Q40583 nicotiana t
20	317.5	11.1	530	10	Q9M535	Q9m535 euphorbia e
21	316.5	11.1	500	10	Q9FVJ6	Q9fvj6 lycopersico
22	313	10.9	521	10	Q9SXM0	Q9sxm0 zea mays (m
23	311	10.9	533	10	Q41152	Q41152 ricinus com
24	307.5	10.7	488	10	Q9FV92	Q9fv92 solanum tub
25	306	10.7	523	10	Q9M422	Q9m422 hordeum vul
26	305.5	10.7	501	10	O65929	O65929 daucus caro
27	302.5	10.6	510	10	Q9SP15	Q9sp15 asarina bar
28	301.5	10.5	429	10	Q40167	Q40167 lycopersico
29	299.5	10.5	505	10	Q9SQK4	Q9sqk4 vitis vinif
30	296.5	10.4	510	10	Q9M3R4	Q9m3r4 arabidopsis
31	294.5	10.3	510	10	Q9FE59	Q9fe59 arabidopsis
32	294	10.3	510	10	Q40938	Q40938 plantago ma
33	294	10.3	523	10	Q39438	Q39438 beta vulgar
34	292.5	10.2	512	10	Q9C8X2	Q9c8x2 arabidopsis
35	292	10.2	537	10	O49838	O49838 oryza sativ
36	292	10.2	538	10	Q9LKH3	Q9lkh3 oryza sativ
37	291.5	10.2	509	10	Q9FNY9	Q9fny9 arabidopsis
38	291	10.2	618	4	Q9ULU3	Q9ulu3 homo sapien
39	287	10.0	539	10	Q9SM24	Q9sm24 beta vulgar
40	286	10.0	474	10	O04516	O04516 arabidopsis
41	284.5	9.9	334	10	O65883	O65883 ricinus com
42	281.5	9.8	491	10	Q9C6H8	Q9c6h8 arabidopsis
43	281	9.8	491	10	Q9FG00	Q9fg00 arabidopsis
44	280.5	9.8	492	10	Q9FIX9	Q9fix9 arabidopsis
45	279.5	9.8	513	10	Q39232	Q39232 arabidopsis

ALIGNMENTS

RESULT 1
Q9VSV1
ID Q9VSV1 PRELIMINARY; PRT; 599 AA.
AC Q9VSV1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CG4484 PROTEIN.
GN CG4484.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

```
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003552; AAF50310.1; -.
DR   FlyBase; FBgn0035968; CG4484.
SQ   SEQUENCE   599 AA;  66057 MW;  C5381D334CFBF2EB CRC64;

  Query Match             18.2%;  Score 519.5;  DB 5;  Length 599;
  Best Local Similarity   23.7%;  Pred. No. 2.2e-28;
  Matches  150;  Conservative  100;  Mismatches  201;  Indels  183;  Gaps   14;

Qy      12 RHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLL 71
           |  : ::  :: :   :|   ||  ::| |:||::||: | |:|  |: |:::    |||
Db      46 RKTRFEMFRLSAIAMAIEFAYAAETSFVSPILLQIGVDHKHMSMTWGLSPLIGFFMSPLL 105

Qy      72 GSASDHWRGRYGRRRPFIWALSLGILLSLFLIP-------------------------R 105
           || ||  : |:|||| |  || ||:  | |:|
Db     106 GSISDRCKLRWGRRRPIISILSFGIMCGLILVPYGKDLGLLLGDAGYTYAESALNFTSSS 165

Qy     106 AGWLAGLLC------PDPRPLELA--LLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHC 157
           | :| |:       |     : |  | |||: |||    | ||    | |:   |:
Db     166 GGSVAALVSGEATTGPSASDYKFAVILTILGMVLLDFDADTCQTPARTYLLDMC-VPEEQ 224

Qy     158 RQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATL-- 215
           :| :::|     || :|| :  :||:|: :  ::|     :| |:|:||  |   |:
Db     225 PKAMTMFALFAGFGGTIGYAIGGVDWETTHIGSFMGGNIPTVFTLVTIIFAVCYLITVTT 284

Qy     216 -------LVAEEAALGP-------------------------------------TEPAEG 231
                 |: ::  | |                                       | :|
Db     285 FREIPLPLIEQDELLRPLSEQAIKKELKKKNNTIYYIQETTQLELQMASDDPKRLEALQG 344

Qy     232 LSAPSLSP----------------HCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLF 275
                 ||              :  |   :   :::             || ::| |
Db     345 SYQNGYSPAVEKQGKSQDLETQSDYDAPVSLKAYLKSIFI----------MPYSMRMLA 393

Qy     276 VAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAIS 335
           :  |  ||  :|: |::|||||| :: | | | | :||  :|: |||  |  |: :
Db     394 LTNLFCWMGHVTYCLYFTDFVGEAVFHGDPTAAPNSEAALNYEAGVRFGCWGMAIYAFSC 453

Qy     336 LVFSLVMDRLVQRFGTRAVYLASVAAFPV---AAGATCLSHSVAVVTASAALTGFTFSAL 392
           ::|| : :|:: |||:|||:: :  : :     |       | | : ||   |  :  :
Db     454 SIYSLSVTKLMKWFGTKAVYISGMIYYGIGMLVLGLWPTKWGVLVFSTSA---GILYGTI 510

Qy     393 QILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSG 452
             :|: | : ||  |  |  |                          ||
Db     511 FTVPFILVARYH-AKNCFSIK----------------------------NG-------- 532

Qy     453 LLPPPPALCGASACDVSVRVVVGEPTEARVVP---GRGICLDLAILDSAFLLSQVAPSLF 509
                                        :||    ||:  |:||: |   ::|:  ||
Db     533 ---------------------------EIVPLKQARGLGTDVAIISSMVFIAQLIVSLS 564

Qy     510 MGSIVQLSQSVTAYMVSAAGLGLVAIYFATQVVF 543
           :| :|    :  | : ::  |  :|  |  |::
Db     565 VGPLVSWMDTTCAVLYASTFLSFLAAIAAMFVLY 598


RESULT   2
Q9UMX9
ID   Q9UMX9      PRELIMINARY;      PRT;   530 AA.
AC   Q9UMX9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   AIM-1 PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA;
RA   Harada M., Li Y.F., El-Gamil M., Rosenberg S.A., Robbins P.F.;
RT   "Use of an in vitro immunoselected tumor line to identify shared
RT   melanoma antigens recognized by HLA-A*0201-restricted T cells.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF172849; AAD51812.1; -.
SQ   SEQUENCE   530 AA;  58301 MW;  F14A4BACAA8FF31B CRC64;

  Query Match             17.1%;  Score 488.5;  DB 4;  Length 530;
  Best Local Similarity   26.4%;  Pred. No. 2.8e-26;
  Matches  150;  Conservative   78;  Mismatches  233;  Indels  107;  Gaps   12;

Qy       8 SRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVC 67
             |||: |  |          || | | |    || |:|| ||:     ::|  : |:|| :
Db      33 SRLIMHSMAM--------FGREFCYAVEAAYVTPVLLSVGLPSSLYSIVWFLSPILGFLL 84

Qy      68 VPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPR---PLELAL 124
             |::|||||| | |:|||||:|  | : :|: : |       :   |  :||      :::
Db      85 QPVVGSASDHCRSRWGRRRPYILTLGVMMLVGMALYLNGATVVAALIANPRRKLVWAISV 144

Qy     125 LILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWD 184
             ::|| | ||        |::| | |:   |  :    :|      || |||||  ||||
Db     145 TMIGVVLFDFAADFIDGPIKAYLFDVCSHQDK-EKGLHYHALFTGFGGALGYLLGAIDWA 203

Qy     185 TSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPC 244
             |   |||: : :|    |:   |    |    || |  || |:|:  |  :|   |
Db     204 HLELGRLLGTEFQVMFFFSALVLTLCFTVHLCSISEAPL--TEVAKGI-PPQQTPQDPPL 260

Qy     245 RA------------------RLAFRNL-----------GALLPRLHQLCCRMPRTLRRL 274
           :                    || :                |  | :    ||   | |
Db     261 SSDGMYEYGSIEKVKNGYVNPELAMQGAKNKNHAEQTRRAMTLKSLLRALVNMPPHYRYL 320

Qy     275 FVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAI 334
           :: |  | | ::  ||:|||:|: :|:| | :   :     |: || :|  |  :
Db     321 CISHLIGWTAFLSNMLFFTDFMGQIVYRGDPYSAHNSTEFLIYERGVEVGCWGFCINSVF 380

Qy     335 SLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQI 394
           | ::|     ||   | : :|      | :  |   |  :|       :| |   | |
Db     381 SSLYSYFQKVLVSYIGLKGLYFTGYLLFGLGTGFIGLFPNVYSTLVLCSLFGVMSSTLYT 440

Qy     395 LPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLL 454
           :|: | : ||||::    | |                ||  |
Db     441 VPFNLITEYHREEE----KERQQA-------------PGGDP----------------- 465

Qy     455 PPPPALCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIV 514
                     | |||             |:|  :| | |     |:|:     :| :|
Db     466 -----------DNSVR-------------GKG--MDCATLTCMVQLAQILVGGGLGFLV 498

Qy     515 QLSQSVTAYMVSAAGLGLVAIYFATQVV 542
             : :|   :::|: : |:   |    |
Db     499 NTAGTVVVVVITASAVALIGCCFVALFV 526


RESULT   3
O65803
ID   O65803      PRELIMINARY;      PRT;   515 AA.
AC   O65803;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SUCROSE/H+ SYMPORTER.
GN   SUT2.
OS   Daucus carota (Carrot).
```

```
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX   NCBI_TaxID=4039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NAMTAISE; TISSUE=ROOTS;
RA   Shakya R., Sturm A.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR   EMBL; Y16768; CAA76369.1; -.
DR   Mendel; 29262; Dauca;2554;29262.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE   515 AA;  54424 MW;  897B90657C9E243C CRC64;


  Query Match             12.1%;  Score 347.5;  DB 10;  Length 515;
  Best Local Similarity   25.9%;  Pred. No. 1.7e-16;
  Matches  131;  Conservative   80;  Mismatches  215;  Indels   79;  Gaps   17;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|:||  :  |::  |  :: : | :  :|:  |:   :   ||: |::  |::|  ||
Db      34 KLVLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 93

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
           | :  :|||||||| : :  : :|: ||   |: | :       :   :|  : : ::|
Db      94 HCQSSFGRRRPFIASGAGCVAISVILI---GFAADISYKAGDDMSKTLKPRAVTVFVIGF 150

Qy     130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LL 178
            :||    :   |  |||:||   |    | | : |:| :::|  |||          |
Db     151 WILDVANNMLQGPCRALLADLCSGDTRRMRSANAFYSFFMAVGNILGYAAGSYNNLYKLF 210

Qy     179 PAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLS 238
           |     | |   |    : |    : |: :  | |  :| | :  || : |:    |
Db     211 PFS--KTHACDLYCANLKSCFIISIALLIIITVVALSVVRENS--GPPDDADAAEEP--- 263

Qy     239 PHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGE 298
           |   ::     | ||  |  |    || :  | :    :|:|   | || ||::|
Db     264 ----PSSGKIPV--FGELLGALKDL----PRPMLLLLIVTCLNWIAWFPFILFDTDWMGR 313

Qy     299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
           :| |   |  |    : ||:||| |:||| |   :  : |: :: ||:  |   : |
Db     314 EIYGGT--AGQG----KLYDQGVRAGALGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LWG 366

Qy     359 VAAFPVAAGATCL----------SHSV--------AVVTASA----ALTGFTFSALQIL 395
           | :| |              ||         | | | |    ::  |   |    :
Db     367 FVNFILAIGLVMTVVVSKVAQHQREHSANGQLLPPSAGVKAGALSLFSILGIPLSITYSI 426

Qy     396 PYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPG-------APFPNGHVGA 448
           |: |||:|            |    |     :: | | ||            |  | : |
Db     427 PFALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAFVVGAISA 486

Qy     449 GGSGLLP----PPPALCGASACDVS 469
             ||:|     | |:   ||   :|
Db     487 AISGVLAIVLLPKPSKDAASKLSLS 511


RESULT   4
Q9FNR6
ID   Q9FNR6      PRELIMINARY;      PRT;   515 AA.
AC   Q9FNR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   SUCROSE/PROTON SYMPORTER.
GN   SUT2.
OS   Daucus carota (Carrot).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX   NCBI_TaxID=4039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NANTAISE;
RA   Sturm A.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NANTAISE;
RA   Shakya R.;
RL   Thesis (2000), Department of Department of Botany,
RL   University of Basel, Basel, Switzerland.
DR   EMBL; AJ303199; CAC19689.1; -.
SQ   SEQUENCE   515 AA;  54425 MW;  9580978C726523DC CRC64;


  Query Match             12.1%;  Score 347.5;  DB 10;  Length 515;
  Best Local Similarity   25.9%;  Pred. No. 1.7e-16;
  Matches  131;  Conservative   80;  Mismatches  215;  Indels   79;  Gaps   17;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|:||  :  |::  |  :: : | :  :|:  |:   :   ||: |::  |::|  ||
Db      34 KLVLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 93

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
           | :  :|||||||| : :  : :|: ||   |: | :       :   :|  : : ::|
Db      94 HCQSSFGRRRPFIASGAGCVAISVILI---GFAADISYKAGDDMSKTLKPRAVTVFVIGF 150

Qy     130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LL 178
            :||    :   |  |||:||   |    | | : |:| :::|  |||          |
Db     151 WILDVANNMLQGPCRALLADLCSGDTRRMRSANAFYSFFMAVGNILGYAAGSYNNLYKLF 210

Qy     179 PAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLS 238
           |     | |   |    : |    : |: :  | |  :| | :  || : |:    |
Db     211 PFS--KTHACDLYCANLKSCFIISIALLIIITVVALSVVRENS--GPPDDADAAEEP--- 263

Qy     239 PHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGE 298
           |   ::     | ||  |  |    || :  | :    :|:|   | || ||::|
Db     264 ----PSSGKIPV--FGELLGALKDL----PRPMLLLLIVTCLNWIAWFPFILFDTDWMGR 313

Qy     299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
           :| |   |  |    : ||:||| |:||| |   :  : |: :: ||:  |   : |
Db     314 EIYGGT--AGKG----KLYDQGVRAGALGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LWG 366

Qy     359 VAAFPVAAGATCL----------SHSV--------AVVTASA----ALTGFTFSALQIL 395
           | :| |              ||         | | | |    ::  |   |    :
Db     367 FVNFILAIGLVMTVVVSKVAQHQREHSANGQLLPPSAGVKAGALSLFSILGIPLSITYSI 426

Qy     396 PYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPG-------APFPNGHVGA 448
           |: |||:|            |    |     :: | | ||            |  | : |
Db     427 PFALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAFVVGAISA 486

Qy     449 GGSGLLP----PPPALCGASACDVS 469
             ||:|     | |:   ||   :|
Db     487 AISGVLAIVLLPKPSKDAASKLSLS 511


RESULT   5
Q9SQK6
ID   Q9SQK6      PRELIMINARY;      PRT;   501 AA.
AC   Q9SQK6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE SUCROSE TRANSPORTER.
GN   VVSUC11.
OS   Vitis vinifera (Grape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;
OC   Vitis.
OX   NCBI_TaxID=29760;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=SHIRAZ; TISSUE=BERRY;
RA   Davies C., Wolf T., Robinson S.P.;
RT   "Three putative sucrose transporters are differentially expressed in
RT   grapevine tissues.";
RL   Plant Sci. 147:93-100(1999).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR   EMBL; AF021808; AAF08329.1; -.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE   501 AA;  53938 MW;  4D0D4DE2EF2F4BA8 CRC64;


  Query Match             12.1%;  Score 346.5;  DB 10;  Length 501;
  Best Local Similarity   27.0%;  Pred. No. 2e-16;
  Matches  153;  Conservative   83;  Mismatches  197;  Indels  133;  Gaps   22;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|| |  :  |::  |  :: : | : |:|:  : :::  ||: ||:  ||:|  ||
Db      29 RLLRVASVACGIQFGWALQLSLLTPYVQELGIPHAWSSIIWLCGPLSGLLVQPLVGHLSD 88

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLL--CPDPRPLELALLILGVGLLDF 134
           |:|||||| | :  |:::: :|  :  : |||    | ||  :|  ::|  |||
Db      89 RCNSRFGRRRPFIVAGATSIVVAVLIIGFSADIGGLLGDGADRRPRAVATFVVGFWLLDV 148

Qy     135 CGQVCFTPLEALLSDLFRDPDH--CRQAYSVYAFMISLGGCLGYLLPA------IDWDTS 186
             |  | |||:|| : ||  | | : ::  |::|  ||:  :      | | ||
Db     149 ANNVTQGPCRALLADL-TEKDHRRTRVANAYFSLFIAVGNVLGFATGSYSGWFRIFWFTS 207

Qy     187 ALAPYLGTQEECL----FGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCC 242
             :     :|      || :||   :| |  ::  ||          ||: | | |
Db     208 TSS----CNADCANLKSAFLLDIIF---IAITTYISITAA-----QELPLSSSSRSTHIS 255

Qy     243 -----PCRARLAFRNLGALLPRLHQLCCRMPRTLRR--------LFVAELCSWMALMTFT 289
                |: ||         | :|      |||         |||  | :|: |: |
Db     256 EEMAESTHAQEAF---------LWELF----GTLRYLSGSIWIILFVTAL-TWIGLLPFL 301

Qy     290 LFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRF 349
           || ||::|  :| | |         ::|: |||||:||| |   :  : |::|::| :::
Db     302 LFDTDWMGREIYGGKP------NEGQNYNTGVRMGALGLMLNSVVLGITSVLMEKLCRKW 355

Qy     350 GTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQV 409
           |   |:                        | |    |  : ||    | : | :
Db     356 GAGFVW----------------------GLSNILMSLCFLLMLILS---AVVKHMD--- 386

Qy     410 FLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVS 469
                                  ||   |    |:| | |          ||   |   :
Db     387 --------------------FLGHDLP----PSGVVIA----------ALIVFSILGIP 411

Qy     470 VRVVVGEP---TEARVVP-GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL-----SQSV 520
           : :    |      |:  | |  | : :|: | :: ||  ||  |   ||      | |:
Db     412 LAITYSVPYALISTRIESLGLGQGLSMGVLNLAIVIPQVIVSLGSGPWDQLFGGGNSPSL 471

Qy     521 TAYMVSAAGLGLVAIYFATQVVFDKS 546
             |:|   |||||    :   |||
Db     472 AVAAVAAFASGLVAILAIPRSSADKS 497


RESULT   6
Q9SLN7
ID   Q9SLN7      PRELIMINARY;      PRT;   515 AA.
AC   Q9SLN7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SUCROSE TRANSPORTER PROTEIN.
GN   CSUT.
OS   Daucus carota (Carrot).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX   NCBI_TaxID=4039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ROOT;
RA   Wu N., Diao F., Zhang L., Huang M.;
RT   "Sucrose transporter protein.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR   EMBL; AB036758; BAA89458.1; -.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE   515 AA;  54523 MW;  2B548D1D9DFF51AC CRC64;


  Query Match             12.1%;  Score 346.5;  DB 10;  Length 515;
  Best Local Similarity   27.2%;  Pred. No. 2e-16;
  Matches  137;  Conservative   76;  Mismatches  215;  Indels   75;  Gaps   19;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|:||  :  |::  |  :: : | :  :|:  |:  :  ||: |::  |::|  ||
Db      34 KLVLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 93

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRA---GWLAG-LLCPDPRPLELALLILGVGLL 132
           | :  :|||||| : :  : :|: ||  |   |: ||  :    :|  : : ::|  :|
Db      94 HCQSSFGRRRPFIASGAGCVAISVILIGFAADIGYKAGDDMSKTLKPRAVTVFVIGFWIL 153

Qy     133 DFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LLPAI 181
           |    :  |  |||:||   |    | | : |:| :::|  |||          | |
Db     154 DVANNMLQGPCRALLADLCSGDTRRMRSANAFYSFFMAVGNILGYAAGSYNNLYKLFPFS 213

Qy     182 DWDTSALAPYLGTQEEC-LFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
             | |   |    : | :  :  || :| ||  | |  |    | : |:    |
Db     214 --KTHACDLYCANLKSCFIISIALLIIITVVA--LSVVREKQWSP-DDADAADEP----- 263

Qy     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
           |   ::     | ||  |  |    || :  | :    :|:|   | || ||::|  :
Db     264 --PSSGKIPV--FGELLGALKDL----PRPMLLLLIVTCLNWIAWFPFILFDTDWMGREI 315

Qy     301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
           | |   |  |    : ||:||| ||||| |   :  : |: :: ||:  |   : |  |
Db     316 YGGT--AGKG----KLYDQGVRAGSLGLLLLNSVVLGLTSIAVEYLVRGVGGVKI-LWGVV 368

Qy     361 AFPVAAGATCL-----------SHSV--------AVVTASA----ALTGFTFSALQILPY 397
           | :| |                 ||          | | |    :: |   |    :|:
Db     369 NFILAIGLVMTVVVSKVAQHQREHSANGQLLPPSAGVKAGALSLFSILGIPLSITYSIPF 428

Qy     398 TLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPG-------APFPNGHVGAGG 450
           |||:|              |    |      :: | | ||            |  | : |
Db     429 ALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAFVVGAISAAI 488

Qy     451 SGLLP----PPPALCGASACDVS 469
           ||:|      | |:   ||   :|
Db     489 SGVLAIVLLPKPSKDAASKLSLS 511


RESULT   7
Q9SQK5
ID   Q9SQK5      PRELIMINARY;      PRT;   612 AA.
AC   Q9SQK5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   PUTATIVE SUCROSE TRANSPORTER.
GN   VVSUC12.
OS   Vitis vinifera (Grape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;
OC   Vitis.
OX   NCBI_TaxID=29760;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SHIRAZ; TISSUE=BERRY;
```

RA Davies C., Wolf T., Robinson S.P.;
RT "Three putative sucrose transporters are differentially expressed in
RT grapevine tissues.";
RL Plant Sci. 147:93-100(1999).
DR EMBL; AF021809; AAF08330.1; -.
SQ SEQUENCE 612 AA; 65711 MW; 8FC3FBD6AB439078 CRC64;

Query Match 12.1%; Score 346; DB 10; Length 612;
Best Local Similarity 25.4%; Pred. No. 2.7e-16;
Matches 125; Conservative 75; Mismatches 178; Indels 114; Gaps 17;

```
Qy      18 LLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASDH 77
Db      66 LILSCMIAAGVQFGWALQLSLLTPYIQTLGIEHAFSSFIWLCGPITGLVVQPCVGIWSDK 125

Qy      78 WRGRYGRRRPFIWALSLGILLSLFLI---PRAGWLAGLLCPDPRPLE------LALLILG 128
Db     126 CSSKYGRRRPFILAGSLMISVAVTIIGFSADIGYLLGDTNMDCRKFKGTRTWAAIIFVLG 185

Qy     129 VGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAID----WD 184
Db     186 FWMLDLANNTVQGPARALLADL-SGPDQRNSANAIFCSWMAVGNILGFSAGASGHWHRW- 243

Qy     185 TSALAPYLGTQEEC-------LFGLLTLIFLT-CVAATLLVAEEAALGPTEPAE-GLSAP 235
Db     244 ----FPFLLNKACCEACGNLKAAFLIAVVFLTLCTLVTLYFAEEVPLMAYQPHHLSDSAP 299

Qy     236 SL-SPHCCPCRARLAFRNL---------------------------------------- 253
Db     300 LLDNPQ------QIGFDNSKSKLDMSAVDNATGNNPESSYEINKNAKHLTPIVQEQNESF 353

Qy     254 ----GALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEP 309
Db     354 SDGPGAVLVNLLTSLRHLPPAMHSVLLVMALSWLSWFPFFLFDTDWMGREVYHGDPKGD- 412

Qy     310 GTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVY-LASVAAFPVAAGA 368
Db     413 -ESAVKAYDAGVREGAFGLLLNSVDLGISSFLIEPMCQRMGARLVWAMSNFIVFACMAGT 471

Qy     369 TCLS-------------HSV----AVVTASA---ALTGFTFSALQILPYTLASLYHREKQ 408
Db     472 AIISLVSVNEYITEGIQHAIGENRAIKIASLVVFALLGFPLSITYSVPFSITA------- 524

Qy     409 VFLPKYRGDTGG 420
Db     525 ----ELTADTGG 532
```

RESULT 8
Q9S7Z5
ID Q9S7Z5 PRELIMINARY; PRT; 512 AA.
AC Q9S7Z5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SUCROSE TRANSPORTER SUT2B.
GN SUT2B OR SUT2A.
OS Apium graveolens (Celery).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Apium.
OX NCBI_TaxID=4045;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MATURE LEAF;
RA Noiraud N., Lemoine R.;
RT "Sucrose transporters in celery.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AF167416; AAD45391.1; -.
DR EMBL; AF167415; AAD45390.1; -.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 512 AA; 54426 MW; 2637535216FF1ED2 CRC64;

Query Match 11.8%; Score 338; DB 10; Length 512;
Best Local Similarity 25.3%; Pred. No. 7.9e-16;
Matches 140; Conservative 87; Mismatches 198; Indels 128; Gaps 21;

```
Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
Db      33 KLILVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 92

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRA---GWLAG-LLCPDPRPLELALLILGVGLL 132
Db      93 RCQSSFGRRRPFIASGAGCVAISVILIGFAADIGYKAGDDMTKTLKPRAVTGFVIGFWIL 152

Qy     133 DFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LLPAI 181
Db     153 DVANNMLQGPCRALLADLCNGDTRRMRSANAFYSFFMAVGNILGYAAGSYNNLYKLFPFS 212

Qy     182 DWDTSALAPYLGTQEEC-LFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
Db     213 --KTHACDLYCANLKSCFIISIVLLIFITVLALT--VVREKQWSPDEADEEPPSSGKIP- 267

Qy     241 CCPCRARLAFRNL-GALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEG 299
Db     268 --------VFGELFGALKD--------LPRPMLMLLVVTCLNWIAWFPFILFDTDWMGRE 311

Qy     300 LYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASV 359
Db     312 IYGGT--AGQG----KLYDQGVRVGSLGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LWGL 364

Qy     360 AAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTG 419
Db     365 VNFLLAIGLV-----MTVVVSKVA----------QHQRQH---------- 389

Qy     420 GASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTE 479
Db     390 ----------GANGQLLPPSA--GVKAGALSLFSILGIPLS 418

Qy     480 ARV-VP----------GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL--SQSVTAYMV 525
Db     419 ITFSIPFALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAFVV 478

Qy     526 ---SAAGLGLVAI 535
Db     479 GAISAAISGVLAI 491
```

RESULT 9
O04077
ID O04077 PRELIMINARY; PRT; 523 AA.
AC O04077;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SUCROSE TRANSPORT PROTEIN.
GN SUT.
OS Vicia faba (Broad bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Vicia.
OX NCBI_TaxID=3906;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COTYLEDON;
RA Weber H., Borisjuk L., Heim U., Sauer N., Wobus U.;
RL Plant Cell 0:0-0(0).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; Z93774; CAB07811.1; -.

```
DR   Mendel; 16626; Vicfa;2554;16626.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE   523 AA;  55229 MW;  F8EBF170212D191A CRC64;


  Query Match             11.8%;  Score 337.5;  DB 10;  Length 523;
  Best Local Similarity   25.8%;  Pred. No. 8.8e-16;
  Matches  134;  Conservative   76;  Mismatches  185;  Indels  125;  Gaps   20;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           ::::|  :  |::   |  :: : | :  :|:   :   :   ||: |::  |::|  ||
Db      36 KIMVVASIAAGVQFGWALQLSLLTPYVQLLGIHHTWAAYIWLCGPISGMLVQPIVGYHSD 95

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
              |:|||||| | |: : :::|||   |: | |       |    ||  : : ::|
Db      96 RCTSRFGRRRPFIAAGSIAVAIAVFLI---GYAADLGHSFGDSLDQKVRPRAIGIFVVGF 152

Qy     130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID------ 182
           :||    :   |  ||| ||   :    | | : ::| :::|  |||   |
Db     153 WILDVANNMLQGPCRALLGDLCAGNQRKTRNANAFFSFFMAVGNVLGYAAGAYSKLYHVF 212

Qy     183 --WDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLV-AEEAALGPTEPA----EGLSAP 235
             | |   |    : | |  |::  || :| : |:  :| || | :      :| |:
Db     213 PFTKTKACNVYCANLKSCFF--LSIALLTVLATSALIYVKETALTPEKTVVTTEDGGSSG 270

Qy     236 SLSPHCCPCRARL--AFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYT 293
           :    ||  :|  ||: |           | :  | :    :|:|   | || |
Db     271 GM-----PCFGQLSGAFKEL-------------KRPMWILLLVTCLNWIAWFPFLLFDT 311

Qy     294 DFVGEGLYQGVPRAEPGTEARRH-YDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTR 352
           |::|: :|       ||    | || ||| |:||| |   :    || :| | :  |
Db     312 DWMGKEVY-------GGTVGEGHAYDMGVREGALGLMLNSVVLGATSLGVDILARGVG-G 363

Qy     353 AVYLASVAAFPVAAGATCLSHSVAVV-----------------------TASAALTGF 387
           |  :  | :   | ||  :| |                           : ||| |
Db     364 VKRLWGIVNFLL---AICLGLTVLVTKLAQHSRQYAPGTGALGDPLPPSEGIKAGALTLF 420

Qy     388 TFSALQI-----LPYTLASLYHREKQVFLPKYRGDTGGASSEDSL-----------MTS 430
           :  : :     :|: |||::             | ||    ||             |
Db     421 SVLGVPLAITYSIPFALASIF-----------SSTSGAGQGLSLGVLNLAIVIPQMFVS 468

Qy     431 FLPGPKPG-------APFPNGHVGAGGSG-----LLPPPP 458
           | ||           |  | | |  ||     ||| ||
Db     469 VLSGPWDALFGGGNLPAFVVGAVAALASGILSIILLPSPP 508


RESULT  10
Q9SP63
ID   Q9SP63      PRELIMINARY;      PRT;   501 AA.
AC   Q9SP63;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SUCROSE TRANSPORTER.
OS   Vitis vinifera (Grape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;
OC   Vitis.
OX   NCBI_TaxID=29760;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. UGNI BLANC; TISSUE=GRAPE BERRY;
RA   Ageorges A., Delrot S., Romieu C.;
RT   "Sucrose transporter from grape berry.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR   EMBL; AF182445; AAD55269.1; -.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
```

```
SQ   SEQUENCE   501 AA;  54075 MW;  4FD19DFAFE539077 CRC64;


  Query Match             11.8%;  Score 336.5;  DB 10;  Length 501;
  Best Local Similarity   26.9%;  Pred. No. 9.8e-16;
  Matches  152;  Conservative   82;  Mismatches  199;  Indels  133;  Gaps   22;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|| |  :  |::   |  :: : | : |:|:   : :::   ||: ||:  ||:|  ||
Db      29 RLLRVASVACGIQFGWALQLSLLTPYVQELGIPHAWSSIIWLCGPLSGLLVQPLVGHLSD 88

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLL--CPDPRPLELALLILGVGLLDF 134
              |:|||||| | :  |:::: :|  :  : |||    | ||  :|  ::|  |||
Db      89 RCNSRFGRRRPFIVAGATSIVVAVLIIGFSTDIGGLLGDGADRRPRAVATFVVGFWLLDV 148

Qy     135 CGQVCFTPLEALLSDLFRDPDH--CRQAYSVYAFMISLGGCLGYLLPA------IDWDTS 186
           |   |  |||:||  : ||   | | : ::  |::|  ||:   :       | | ||
Db     149 ANNVTQGPCRALLADL-TEKDHRRTRVANAYFSLFIAVGNVLGFATGSYSGWFRIFWFTS 207

Qy     187 ALAPYLGTQEECL----FGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCC 242
           :       :|       || :||   :| |  ::  ||          ||: | | |
Db     208 TSS----CNADCANLKSAFLLDIIF---IAITTYISITAA-----QELPLSSSSRSTHIS 255

Qy     243 -----PCRARLAFRNLGALLPRLHQLCCRMPRTLRR--------LFVAELCSWMALMTFT 289
                 |: ||          | :|      |||          |||  | :|:    |
Db     256 EEMAESTHAQEAF---------LWELF----GTLRYLSGSIWIILFVTAL-TWIGWFPFL 301

Qy     290 LFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRF 349
           || ||::|  :| | |          ::|: |||||:||| |   :  : |::|::| :::
Db     302 LFDTDWMGREIYGGKP------NEGQNYNTGVRMGALGLMLNSVVLGITSVLMEKLCRKW 355

Qy     350 GTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQV 409
           |   |:                        | |    |  : ||    | : | :
Db     356 GAGFVW----------------------GLSNILMSLCFLLMLILS---AVVKHMD--- 386

Qy     410 FLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVS 469
                                 ||    |    |:| | |          ||   |   :
Db     387 --------------------FLGHDLP----PSGVVIA----------ALIVFSILGIP 411

Qy     470 VRVVVGEP---TEARVVP-GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL-----SQSV 520
           : :    |      |:   | |  | : :|: | :: ||  ||  |   ||     | |:
Db     412 LAITYSVPYALISTRIESLGLGQGLSMGVLNLAIVIPQVIVSLGSGPWDQLFGGGNSPSL 471

Qy     521 TAYMVSAAGLGLVAIYFATQVVFDKS 546
             |:|   |||||    :   |||
Db     472 AVAAVAAFASGLVAILAIPRSSADKS 497


RESULT  11
Q9ZTB9
ID   Q9ZTB9      PRELIMINARY;      PRT;   512 AA.
AC   Q9ZTB9;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SUCROSE TRANSPORTER.
GN   SUT1.
OS   Apium graveolens (Celery).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids II; Apiales; Apiaceae; Apium.
OX   NCBI_TaxID=4045;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LEAF;
RA   Noiraud N., Delrot S., Lemoine R.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR   EMBL; AF063400; AAC99332.1; -.
DR   Mendel; 36163; Apigr;2554;36163.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
```

```
KW   Transmembrane.
SQ   SEQUENCE   512 AA;  54520 MW;  4D76A7854A7DF903 CRC64;


  Query Match             11.6%;  Score 333;  DB 10;  Length 512;
  Best Local Similarity   25.0%;  Pred. No. 1.8e-15;
  Matches  139;  Conservative   86;  Mismatches  198;  Indels  132;  Gaps   21;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|:||  :  |::  |  :: : | :  :|:  |:   :   ||: |::  |::|  ||
Db      33 KLILVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 92

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRA---GWLAGLLCPDP-----RPLELALLILG 128
             :  :|||||| : :  : :|: ||  |   |: ||    |      :|  : : ::|
Db      93 RCQSSFGRRRPFIASGAGCVAISVILIGFAADIGYKAG----DDMNKTLKPRAVTVFVIG 148

Qy     129 VGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLG---------YLL 178
           :||    :   |  |||:||   |    | | ::| | :::|  ||          | |
Db     149 FWILDVANNMLQGPCRALLADLCNGDTRRMRSANALYRFFMAVGNILGNAAGSYNNLYKL 208

Qy     179 PAIDWDTSALAPYLGTQEEC-LFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
             | |   |    : | :  :: |||:| :| |  |  |    | |  |   :
Db     209 FPFS-KTHACDLYCANLKSCFIISIVLLIFITVLALT--VVREKQWSPDEADEEPPSSGK 265

Qy     238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
           |              | ||  |  |    || :  |      :|:|   | || ||::|
Db     266 IP------------VFGELLRALKDL----PRPMLMLLAVTCLNWIAWFPFILFDTDWMG 309

Qy     298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLA 357
           :| |   |  |    : ||:|||:||||| |   :  : |: :: ||:  |   : |
Db     310 REIYGGT--AGQG----KLYDQGVRVGSLGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LW 362

Qy     358 SVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGD 417
           :  | :| |       : || :  |                    |: :
Db     363 GLVNFLLAIGLV-----MTVVVSKVA------------------QHQRQH--------- 389

Qy     418 TGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEP 477
                                                | :| | || |  |  |  :|:  ::| |
Db     390 -----------------------------GANGQLLPPSA--GVKAGALSLFSILGIP 416

Qy     478 TEARV-VP-----------GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL--SQSVTAY 523
             :|           | |  | | :|: | :: |:  |:  |    |    :: |:
Db     417 LSITFSIPFALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAF 476

Qy     524 MV---SAAGLGLVAI 535
           :|   |||  |::||
Db     477 VVGAISAAISGVLAI 491


RESULT  12
Q43653
ID   Q43653      PRELIMINARY;      PRT;   516 AA.
AC   Q43653;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SUCROSE TRANSPORT PROTEIN.
GN   SUT1.
OS   Solanum tuberosum (Potato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4113;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. DESIREE; TISSUE=MATURE LEAF;
RX   MEDLINE=94146554; PubMed=8312741;
RA   Riesmeier J., Willmitzer L., Frommer W.B.;
RT   "Potato sucrose transporter expression in minor veins indicates a role
RT   in phloem loading.";
RL   Plant Cell 5:1591-1598(1993).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR   EMBL; X69165; CAA48915.1; -.
DR   Mendel; 16637; Soltu;2554;16637.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE   516 AA;  54830 MW;  4FD06C095E49A377 CRC64;


  Query Match             11.5%;  Score 330;  DB 10;  Length 516;
  Best Local Similarity   24.7%;  Pred. No. 2.9e-15;
  Matches  132;  Conservative   93;  Mismatches  197;  Indels  112;  Gaps   21

Qy       4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
           :||          ::::|  :  |::   |  :: : | :  :|:  || : :   ||:
Db      29 KLW----------KIIVVASIAAGVQFGWALQLSLLTPYVQLLGIPHKFASFIWLCGPIS 78

Qy      64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPD 116
           |::  |::|  ||:   |:|||||| | :  ::::|||   |: | |       |
Db      79 GMIVQPVVGYYSDNCSSRFGRRRPFIAAGAALVMIAVFLI---GFAADLGHASGDTLGKG 135

Qy     117 PRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPD-HCRQAYSVYAFMISLGGCLG 175
           :|  :|: ::|  :||     :   |  |||:||          | | : ::| :::|  ||
Db     136 FKPRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKSGRMRTANAFFSFFMAVGNILG 195

Qy     176 YLLPAIDW--------DTSALAPYLGTQEECLF-GLLTLIFLTCVAATLLVAEE-AALGP 225
           |   :          | |   |    : | |  :  |: || :| ||:   |
Db     196 YAAGSYSHLFKVFPFSKTKACDMYCANLKSCFFIAIFLLLSLTTIALTLVRENELPEKDE 255

Qy     226 TEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMAL 285
           |  | |:            :::: |   | :   | :|    || :  | :    :|:|
Db     256 QEIDEKLAGAG--------KSKVPF--FGEIFGALKEL----PRPMWILLLVTCLNWIAW 301

Qy     286 MTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRL 345
             | |: ||:: : :: |    : |    | || ||| |::|| ||  :    || :: |
Db     302 FPFFLYDTDWMAKEVFGG----QVGD--ARLYDLGVRAGAMGLLLQSVVLGFMSLGVEFL 355

Qy     346 VQRFGTRAVYLASVAAFPVAAGATCLSHSVAVV---------TASAALTGFT-------- 388
           :: |  |  |  :  | :   | ||: :: |            :  | | |
Db     356 GKKIG-GAKRLWGILNFVL---AICLAMTILVTKMAEKSRQHDPAGTLMGPTPGVKIGAL 411

Qy     389 --FSALQI-------LPYTLASLYHREK-------------QVFLPKYRGDTGGASSEDS 426
             |:|| |       :|: |||::   :               : :|:      |   :|
Db     412 LLFAALGIPLAATFSIPFALASIFSSNRGSGQGLSLGVLNLAIVVPQMLVSLVGGPWDDL 471

Qy     427 LMTSFLPGPKPGAPFPNGHVGAGGSG-----LLPPPPALCGASACDVSVRVVVG 475
             |||      |  | | |  |      :|| |||          |   | :|
Db     472 FGGGNLPG------FVVGAVAAAASAVLALTMLPSPPA-------DAKPAVAMG 512


RESULT  13
Q9XHL6
ID   Q9XHL6      PRELIMINARY;      PRT;   524 AA.
AC   Q9XHL6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SUCROSE TRANSPORT PROTEIN SUT1.
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Pisum.
OX   NCBI_TaxID=3888;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tegeder M., Wang X.-D., Frommer W.B., Offler C.E., Patrick J.W.;
RT   "Sucrose transport into developing seeds of Pisum sativum L.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR   EMBL; AF109922; AAD41024.1; -.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
```

```
SQ   SEQUENCE   524 AA;  55279 MW;  B5CE3F880D2C1E6B CRC64;

  Query Match           11.5%;  Score 329.5;  DB 10;  Length 524;
  Best Local Similarity  25.5%;  Pred. No. 3.2e-15;
  Matches  133;  Conservative   74;  Mismatches  187;  Indels  127;  Gaps   19;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           ::::|  :  |::   |  :: : | :  :|:   :   :   ||: |::  |::|  ||
Db      37 KIMVVASIAAGVQFGWALQLSLLTPYVQLLGIHHTWAAYIWLCGPISGMLVQPVVGYHSD 96

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
              |:|||||| | |: : :::|||   |: | |       |    ||  : : ::|
Db      97 RCTSRFGRRRPFIAAGSIAVAIAVFLI---GYAADLGHSFGDNLDKKVRPRAIGIFVVGF 153

Qy     130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID------ 182
            :||    :   |  ||| ||   :    | | : ::| :::|  |||   |
Db     154 WILDVANNMLQGPCRALLGDLCAGNQRKTRNANAFFSFFMAVGNVLGYAAGAYSKLYHVF 213

Qy     183 --WDTSALAPYLGTQEECLFGLLTLIFLTCVAA------TLLVAEEAALGPTEPAEGLSA 234
             | |   |    : | |  : |: :   ||      | |:|:| :      :| |
Db     214 PFTKTEACNVYCANLKSCFFLSIALLTVLATAALIYVKETPLIAEKAVV---TAEDGGSN 270

Qy     235 PSLSPHCCPCRARL--AFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFY 292
           :   ||  :|  ||: |             | :  | :    :|:|   | ||
Db     271 GGM-----PCFGQLSGAFKEL-------------KRPMWILLLVTCLNWIAWFPFLLFD 311

Qy     293 TDFVGEGLYQGVPRAEPGTEARRH-YDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGT 351
           ||::|: :|        ||    | || ||| |:||| |   :    || :| | :  |
Db     312 TDWMGKEVY-------GGTVGEGHAYDMGVRAGALGLMLNSVVLGATSLGVDILARGVG- 363

Qy     352 RAVYLASVAAFPVAAGATCLSHSVAVV------------------------TASAALTG 386
             |  :  | :   | ||  :| |                            : |||
Db     364 GVKRLWGIVNFLL---AICLGLTVLVTKLAQHSRQYAPGTGGLQDPLPPSGGIKAGALTL 420

Qy     387 FTFSALQI-----LPYTLASLYHREKQVFLPKYRGDTGGASSEDSL------------MT 429
           |:   : :     :|: |||::              | ||    ||
Db     421 FSVLGIPLAITYSIPFALASIF------------SSTSGAGQGLSLGVLNLAIVIPQMFV 468

Qy     430 SFLPGPKPG-------APFPNGHVGAGGSG-----LLPPPP 458
           | | ||           |  | | |  ||     ||| ||
Db     469 SVLSGPWDALFGGGNLPAFVVGAVAALASGILSMILLPSPP 509


RESULT  14
O80605
ID   O80605      PRELIMINARY;      PRT;   594 AA.
AC   O80605;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PUTATIVE SUCROSE/H+ SYMPORTER (SUCROSE TRANSPORTER).
GN   T17M13.3 OR SUC3.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
RA   Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
RT   "Arabidopsis thaliana chromosome II BAC T17M13 genomic sequence.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Meyer S., Truernit E., Huemmer C., Besenbeck R., Stadler R., Sauer N.;
RT   "AtSUC3: an unusual sucrose transporter from Arabidopsis.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004138; AAC32907.1; -.
DR   EMBL; AJ289165; CAB92307.1; -.
SQ   SEQUENCE   594 AA;  63972 MW;  DF075A3601299DA0 CRC64;

  Query Match           11.5%;  Score 328.5;  DB 10;  Length 594;
  Best Local Similarity  24.1%;  Pred. No. 4.3e-15;
  Matches  120;  Conservative   72;  Mismatches  174;  Indels  131;  Gaps   18;

Qy      22 NLLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGS 73
           :|:|  |   :|||: :        : | :  :|:   | : :   ||: |||  | :|
Db      59 SLVTLVLSCTVAAGVQFGWALQLSLLTPYIQTLGISHAFSSFIWLCGPITGLVVQPFVGI 118

Qy      74 ASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPD----------PRPLE 121
           ||    :||||||||   |  | ::: :|   |: |  | |  |           |
Db     119 WSDKCTSKYGRRRPFILVGSFMISIAVIII---GFSADIGYLLGDSKEHCSTFKGTRTRA 175

Qy     122 LALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAI 181
             : |:|  |||        |  |||:||   ||    | :|:   :::|  ||:   |
Db     176 AVVFIIGFWLLDLANNTVQGPARALLADL-SGPDQRNTANAVFCLWMAIGNILGFSAGAS 234

Qy     182 ----DWDTSALAPYLGTQEECLFG--------LLTLIFLT-CVAATLLVAEEAALGPTEP 228
               :|      |:| |   |            || ::||| |   |:  |:|      :|
Db     235 GKWQEW-----FPFL-TSRACCAACGNLKAAFLLAVVFLTICTLVTIYFAKEIPFTSNKP 288

Qy     229 AE-GLSAPSLSPHCCPCRARLAFRNL---------------------------------- 253
              |||  |           |  : |
Db     289 TRIQDSAPLLDD--------LQSKGLEHSKLNNGTANGIKYERVERDTDEQFGNSENEHQ 340

Qy     254 --------GALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVP 305
                  |::|  |       :|  :  : :    :|::   | || ||::|  :| | |
Db     341 DETYVDGPGSVLVNLLTSLRHLPPAMHSVLIVMALTWLSWFPFFLFDTDWMGREVYHGDP 400

Qy     306 RAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVY-LASVAAFPV 364
           :  :     ||:||| |:||| |   :  : | ::: : || | | |: |::    |
Db     401 TGD--SLHMELYDQGVREGALGLLLNSVVLGISSFLIEPMCQRMGARVVWALSNFTVFAC 458

Qy     365 AAGATCLS--------------------HSVAVVTASAALTGFTFSALQILPYTLASLY 403
           ||   :|                     : ||:    || ||  :    :|::: :
Db     459 MAGTAVISLMSLSDDKNGIEYIMRGNETTRTAAVIV--FALLGFPLAITYSVPFSVTA-- 514

Qy     404 HREKQVFLPKYRGDTGG 420
                  :   |:||
Db     515 ---------EVTADSGG 522


RESULT  15
Q9FVL6
ID   Q9FVL6      PRELIMINARY;      PRT;   604 AA.
AC   Q9FVL6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   SUCROSE TRANSPORTER-LIKE PROTEIN.
GN   SUT2.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4081;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20359188; PubMed=10899981;
RA   Barker L., Kuhn C., Weise A., Schulz A., Gebhardt C., Hirner B.,
RA   Hellmann H., Schulze W., Ward J.M., Frommer W.B.;
RT   "SUT2, a putative sucrose sensor in sieve elements.";
RL   Plant Cell 12:1153-1164(2000).
DR   EMBL; AF166498; AAG12987.1; -.
SQ   SEQUENCE   604 AA;  65082 MW;  3403349B1692B2AF CRC64;

  Query Match           11.4%;  Score 325.5;  DB 10;  Length 604;
  Best Local Similarity  22.3%;  Pred. No. 7e-15;
  Matches  133;  Conservative   89;  Mismatches  210;  Indels  165;  Gaps   21;

Qy      22 NLLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGS 73
```

```
          :|||  |   :|||: :        : | :  :|:|  | : :   ||: |||  | :|
Db     62 SLLTLILSCTVAAGVQFGWALQLSLLTPYIQTLGIEHAFSSFIWLCGPITGLVVQPCVGI 121

Qy     74 ASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPD----------PRPLE 121
           ||    :|||||||||:  :: | ::: :|   |: |  | |  |           |
Db    122 WSDKCHSKYGRRRPFIFIGAVMISIAVIII---GFSADIGYLLGDTKEHCSTFKGTRSRA 178

Qy    122 LALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAI 181
            : ::|  :||        |  |||:||   ||    | :|:   :::|  ||:   |
Db    179 AIVFVVGFWMLDLANNTVQGPARALLADL-SGPDQRNTANAVFCSWMAVGNILGFSAGAS 237

Qy    182 D----WDTSALAPYLGTQEEC-------LFGLLTLIFLT-CVAATLLVAEEAALGPTE-P 228
            |      |:|  :  |          |: ::||| |   ||  | |  | | :
Db    238 GGWHRW-----FPFLTNRACCEPCGNLKAAFLVAVVFLTLCTLVTLYFANEVPLSPKQYK 292

Qy    229 AEGLSAPSL-SPH-----CCPCRARLAFRNL---------------------------- 253
           ||| | ||          :  | :||
Db    293 RMSDSAPLLDSPQNTGFDLSQSKRELQYRNSVANNESEMGHVADNSPKNEEQRPDKDQGD 352

Qy    254 ------GALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRA 307
                ||:|  |      :|  :  : :    :|::   | || ||::|  :| | |:
Db    353 SFADSPGAVLVNLLTSLRHLPPAMHSVLIVMALTWLSWFPFFLFDTDWMGREVYHGDPKG 412

Qy    308 EPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS--------- 358
          |   :    |::|||  |: || |   :  | | ::: : :  |:| |:  |
Db    413 E--ADEVNAYNQGVREGAFGLLLNSVVLGVSSFLIEPMCKWIGSRLVWAVSNFIVFVCMA 470

Qy    359 ----VAAFPVAAGATCLSHSVAVV--TASAALTGFTFSALQI-----LPYTLASLYHREK 407
              ::   ::|   : | :     |  |||  |:   : :     :|::: :
Db    471 CTAIISVVSISANTQGVQHVIGATRSTQIAALVVFSLLGIPLAVTYSVPFSITA------ 524

Qy    408 QVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACD 467
            :  | |                              || ||
Db    525 -----ELTADAG----------------------------GGQGL-------------- 536

Qy    468 VSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYM 524
           :|    | |||   : |     |: |    : |:  : |:  |:  : | :
Db    537 -----AIGVLNLAIVVPQMVVSLGAGPWDALFGGGNI-PAFALASLAALAAGIFAML 587
```

Search completed: February 26, 2002, 07:57:37
Job time: 137 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 07:56:35 ; Search time 13.31 Seconds
(without alignments)
1523.340 Million cell updates/sec

Title: US-09-605-783A-113
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	324.5	11.3	525	1	STP_SPIOL	Q03411 spinacia ol
2	142	5.0	544	1	YD74_SYNY3	P74168 synechocyst
3	125.5	4.4	399	1	TCR1_ECOLI	P02982 escherichia
4	121.5	4.2	503	1	PUR8_STRLP	P42670 streptomyce
5	119.5	4.2	491	1	AMPG_ECOLI	P36670 escherichia
6	118.5	4.1	496	1	SMVA_SALTY	P37594 salmonella
7	117	4.1	440	1	YHJE_ECOLI	P37643 escherichia
8	111.5	3.9	640	1	Y051_MYCTU	Q10880 mycobacteri
9	111	3.9	473	1	PHDK_NOCSK	O24723 nocardioide
10	110.5	3.9	680	1	CA1A_HUMAN	Q03692 homo sapien
11	110	3.8	368	1	GALT_HUMAN	O60755 homo sapien
12	110	3.8	606	1	NQOC_THETH	Q56227 thermus aqu
13	109	3.8	477	1	YPUM_RHOCA	P26176 rhodobacter
14	107	3.7	419	1	CMLA_PSEAE	P32482 pseudomonas
15	105	3.7	448	1	YJ94_YEAST	P47159 saccharomyc
16	104	3.6	476	1	MELB_SALTY	P30878 salmonella
17	103	3.6	465	1	FTSW_MYCLE	Q50186 mycobacteri
18	102.5	3.6	461	1	PUCC_RHOCA	P23462 rhodobacter
19	102	3.6	438	1	SHIA_ECOLI	P76350 escherichia
20	101.5	3.5	428	1	YXIO_BACSU	P42306 bacillus su
21	101	3.5	437	1	BRAZ_PSEAE	P25185 pseudomonas
22	101	3.5	471	1	MELB_ENTAE	O07366 enterobacte
23	101	3.5	471	1	MELB_KLEPN	Q02581 klebsiella
24	100	3.5	404	1	Y4XM_RHISN	P55705 rhizobium s
25	99	3.5	481	1	LMRA_STRLN	P46104 streptomyce
26	98.5	3.4	613	1	NUOL_ECOLI	P33607 escherichia
27	98	3.4	379	1	CYB_SORCI	O79451 sorex ciner
28	98	3.4	496	1	GTR3_CHICK	P28568 gallus gall
29	97.5	3.4	385	1	PI2R_BOVIN	P79393 bos taurus
30	97.5	3.4	433	1	DCUA_WOLSU	O34245 wolinella s
31	97.5	3.4	1103	1	CYGD_HUMAN	Q02846 homo sapien
32	97	3.4	336	1	CYB_SORGA	O79996 sorex grana
33	97	3.4	336	1	CYB_SORHA	O79452 sorex hayde
34	97	3.4	336	1	CYB_SORMO	O79969 sorex monti
35	97	3.4	336	1	CYB_SORVA	O80019 sorex vagra
36	97	3.4	370	1	GALT_RAT	O88626 rattus norv
37	97	3.4	427	1	RBTT_KLEPN	O52717 klebsiella
38	96.5	3.4	400	1	TCR8_PASMU	P51564 pasteurella
39	96.5	3.4	473	1	GLF_ZYMMO	P21906 zymomonas m
40	96.5	3.4	485	1	YIHO_ECOLI	P32136 escherichia
41	96	3.4	336	1	CYB_SORCO	O79448 sorex coron
42	96	3.4	405	1	TCR5_ECOLI	Q07282 escherichia
43	96	3.4	460	1	YAGG_ECOLI	P75683 escherichia
44	96	3.4	469	1	MELB_ECOLI	P02921 escherichia
45	95.5	3.3	424	1	CYAA_STIAU	P40137 stigmatella

ALIGNMENTS

```
RESULT 1
STP_SPIOL
ID   STP_SPIOL      STANDARD;      PRT;   525 AA.
AC   Q03411;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   SUCROSE TRANSPORT PROTEIN (SUCROSE PERMEASE) (SUCROSE-PROTON
DE   SYMPORTER).
OS   Spinacia oleracea (Spinach).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
OX   NCBI_TaxID=3562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Leaf;
RX   MEDLINE=93099843; PubMed=1464305;
RA   Riesmeier J.W., Willmitzer L., Frommer W.B.;
RT   "Isolation and characterization of a sucrose carrier cDNA from
RT   spinach by functional expression in yeast.";
RL   EMBO J. 11:4705-4713(1992).
CC   -!- FUNCTION: RESPONSIBLE FOR THE TRANSPORT OF SUCROSE INTO THE
CC       CELL, WITH THE CONCOMITANT EXPORT OF A PROTON (SYMPORT SYSTEM).
CC       CAN ALSO TRANSPORT MALTOSE AT A LESSER RATE.
CC   -!- PATHWAY: SUCROSE METABOLIC PATHWAY.
CC   -!- SUBCELLULAR LOCATION: INNER MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboratio
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no wa
CC   modified and this statement is not removed. Usage by and for commercia
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X67125; CAA47604.1; -.
DR   PIR; S28052; S28052.
DR   InterPro; IPR003662; sub_trnsportr.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
KW   Transmembrane; Transport; Sugar transport; Symport; Glycoprotein.
FT   DOMAIN        1     40       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     41     60       POTENTIAL.
FT   TRANSMEM     70     88       POTENTIAL.
FT   TRANSMEM    108    127       POTENTIAL.
FT   TRANSMEM    146    165       POTENTIAL.
FT   TRANSMEM    185    204       POTENTIAL.
FT   TRANSMEM    230    249       POTENTIAL.
FT   TRANSMEM    296    315       POTENTIAL.
FT   TRANSMEM    341    360       POTENTIAL.
FT   TRANSMEM    373    392       POTENTIAL.
FT   TRANSMEM    419    438       POTENTIAL.
FT   TRANSMEM    441    461       POTENTIAL.
```

```
FT   TRANSMEM    490    509       POTENTIAL.
FT   DOMAIN      510    525       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    270    270       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    271    271       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   525 AA;  54992 MW;  018347A4D2CC1CC6 CRC64;
```

```
  Query Match             11.3%;  Score 324.5;  DB 1;  Length 525;
  Best Local Similarity   23.2%;  Pred. No. 4.5e-16;
  Matches  129;  Conservative   91;  Mismatches  196;  Indels  141;  Gaps   20;

Qy      23 LLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSA 74
           |   ||   :|||: :          : | :  :|:   :   :   ||: |::  ||:|
Db      35 LKKLGLVASVAAGVQFGWALQLSLLTPYVQLLGIPHTWAAYIWLCGPISGMIVQPLVGYY 94

Qy      75 SDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPDP-----RPLELALLIL 127
           ||    |:|||||| | :  : ::: ||   |: |  |    ||     :|  :|: ::
Db      95 SDRCTSRFGRRRPFIAAGAALVAVAVGLI---GFAADIGAASGDPTGNVAKPRAIAVFVV 151

Qy     128 GVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID---- 182
           |  :||         |  |||:|:         | | : ::| ::||   ||   :
Db     152 GFWILDVANNTLQGPCRALLADMAAGSQTKTRYANAFFSFFMALGNIGGYAAGSYSRLYT 211

Qy     183 ----WDTSALAPYLGTQEECLFGLLT-LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
               |:|   |    : | |  :| || || :| :::   :   :   :  | |  :
Db     212 VFPFTKTAACDVYCANLKSCFFISITLLIVLTILALSVVKERQITIDEIQEEEDLKNRNN 271

Qy     238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
           |  |    ||| |   | |:  |  |    |: :  | :    :|:|   | || ||::|
Db     272 SSGC----ARLPF--FGQLIGALKDL----PKPMLILLLVTALNWIAWFPFLLFDTDWMG 321

Qy     298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG--TRAVY 355
           : :| |           : ||:|| |:||| :   :  | || :: | :  |   |
Db     322 KEVYGGT------VGEGKLYDQGVHAGALGLMINSVVLGVMSLSIEGLARMVGGAKRLWG 375

Qy     356 LASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYR 415
           : ::        | ||: :| :|| ||                 : |:    :
Db     376 IVNIIL------AVCLAMTV-LVTKSAE------------------HFRDSHHIM---- 405

Qy     416 GDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVG 475
                                      || : |||||  |     :::  |:|
Db     406 ---------------------------------GSAVPPPPPA--GVKGGALAIFAVLG 429

Qy     476 EPTEARV-VP----------GRGICLDLAILDSAFLLSQV----------------AP 506
           |      :|           | |  | | :|: | :: |:                 |
Db     430 IPLAITFSIPFALASIFSASSGSGQGLSLGVLNLAIVVPQMFVSVTSGPWDAMFGGGNLP 489

Qy     507 SLFMGSIVQLSQSVTAY 523
           :  :|::   : :| ::
Db     490 AFVVGAVAATASAVLSF 506
```

```
RESULT   2
YD74_SYNY3
ID   YD74_SYNY3     STANDARD;      PRT;   544 AA.
AC   P74168;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL SYMPORTER SLL1374.
GN   SLL1374.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:GALACTOSIDE SYMPORTER FAMILY
CC       (SGF).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90912; BAA18257.1; -.
DR   InterPro; IPR001927; Na_galact_symp.
DR   Pfam; PF01236; Na_Galacto_symp; 1.
DR   PROSITE; PS00872; NA_GALACTOSIDE_SYMP; FALSE_NEG.
KW   Hypothetical protein; Transport; Transmembrane; Symport;
KW   Complete proteome.
FT   TRANSMEM     31     51       POTENTIAL.
FT   TRANSMEM     52     72       POTENTIAL.
FT   TRANSMEM     84    104       POTENTIAL.
FT   TRANSMEM    116    136       POTENTIAL.
FT   TRANSMEM    162    182       POTENTIAL.
FT   TRANSMEM    191    211       POTENTIAL.
FT   TRANSMEM    230    250       POTENTIAL.
FT   TRANSMEM    257    277       POTENTIAL.
FT   TRANSMEM    318    338       POTENTIAL.
FT   TRANSMEM    356    376       POTENTIAL.
FT   TRANSMEM    383    403       POTENTIAL.
FT   TRANSMEM    407    427       POTENTIAL.
FT   TRANSMEM    450    470       POTENTIAL.
FT   TRANSMEM    501    521       POTENTIAL.
SQ   SEQUENCE   544 AA;  60064 MW;  C91D0EDFF32277EE CRC64;
```

```
  Query Match              5.0%;  Score 142;  DB 1;  Length 544;
  Best Local Similarity   22.3%;  Pred. No. 0.0049;
  Matches  105;  Conservative   60;  Mismatches  179;  Indels  126;  Gaps   20;

Qy      16 AQLLLVNLLTFGLEVCLAAG----------ITYVPPLLLEV-GVEEKFMTMVLGIGPVL 63
           :| |    | |  ::   ||           : |:   | :| |:       || || :
Db       2 SQSLSAEKLHFTTKLAYGAGDFGPAITANILVFYLLFFLTDVAGIPAALAGSVLMIGKIF 61

Qy      64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPLELA 123
             :  |::|  ||  | |:||| |::    || ::   |   | ||       |    :
Db      62 DAINDPIIGLLSDRTRSRWGRRLPWM----LGGMIPFALFYTAQWLIPHFSDDRLTNQWG 117

Qy     124 LLILGVGLLDFCGQVCFT----PLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLP 179
           | |  | :      :|:|    |  ||  :| :: :  |   : : |  |:|| :  |:
Db     118 LFIYYVAIA-MAFNLCYTTVNLPYTALTPELTQNYNE-RTRLNSFRFAFSIGGSILSLIL 175

Qy     180 AIDWDTSALAPYLGTQEECLFGLL-TLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLS 238
           |      :|  |  : :  || |  :| :  ::| |  |         ||   : :|||
Db     176 YI-----LIAAGLPDRPQQQFGELGVMISVLSISALLWSALRLQEKGKEP---ILSPSL- 226

Qy     239 PHCCPCRARLA----------------`-FRNLGAL------------------------ 256
                 | |||                  |  ||
Db     227 ------RRRLAPLLMAAGITLILLAIAKSFNLLGGSGFDYISFFLILLGLIWGGFGFTLR 280

Qy     257 ----------------------LPRLHQLCCRMPRTLRR-LFV--AELCSWMALMTFTLF 291
                                 || | ||  ::  : |  |||    ||||:|:
Db     281 DSAVEEHLQKLENSPSPGVTENLPLLKQL--KIAFSNRAFLFVIGIYLCSWLAVQLTASI 338

Qy     292 YTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGT 351
             ||    : |:   : ||               : | :|   :||   |   | |
Db     339 LVYFVVS--WMGLNEQQSGT-------------IALAVQ-GTALVMLFVWQALAQFLDK 381

Qy     352 RAVY-LASVAAFPVAAGATCLS-HSVAVVTASAALTGFTFSALQILPYTL 399
           : :| | |:       ||   :    ||::   |   |   |   ::|:::
Db     382 KVIYFLGSMVWMGAEAGLWLVQPGQVALLYTLAIFAGVGVSVAYLIPWSM 431
```

```
RESULT   3
TCR1_ECOLI
ID   TCR1_ECOLI     STANDARD;      PRT;   399 AA.
AC   P02982;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   TETRACYCLINE RESISTANCE PROTEIN, CLASS A (TETA(A)).
GN   TETA.
OS   Escherichia coli.
OG   Plasmid RP1.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=Tn1721;
RX   MEDLINE=92192465; PubMed=1312499;
RA   Allmeier H., Cresnar B., Greck M., Schmitt R.;
RT   "Complete nucleotide sequence of Tn1721: gene organization and a
RT   novel gene product with features of a chemotaxis protein.";
RL   Gene 111:11-20(1992).
RN   [2]
RP   SEQUENCE OF 85-399 FROM N.A.
RC   STRAIN=DH1;
RA   Fletcher J.N., Hart C.A., Batt R.M., Saunders J.R.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=Tn1721; PLASMID=RP1;
RX   MEDLINE=83299270; PubMed=6310527;
RA   Waters S.H., Rogowsky P., Grinsted J., Altenbuchner J., Schmitt R.;
RT   "The tetracycline resistance determinants of RP1 and Tn1721:
RT   nucleotide sequence analysis.";
RL   Nucleic Acids Res. 11:6089-6105(1983).
RN   [4]
RP   TOPOLOGY.
RX   MEDLINE=92388137; PubMed=1517220;
RA   Allard J.D., Bertrand K.P.;
RT   "Membrane topology of the pBR322 tetracycline resistance protein.
RT   TetA-PhoA gene fusions and implications for the mechanism of TetA
RT   membrane insertion.";
RL   J. Biol. Chem. 267:17809-17819(1992).
CC   -!- FUNCTION: RESISTANCE TO TETRACYCLINE BY AN ACTIVE TETRACYCLINE
CC       EFFLUX. THIS IS AN ENERGY-DEPENDENT PROCESS THAT DECREASES THE
CC       ACCUMULATION OF THE ANTIBIOTIC IN WHOLE CELLS. THIS PROTEIN
CC       FUNCTIONS AS A METAL-TETRACYCLINE/H+ ANTIPORTER.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC       AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X61367; CAA43643.1; -.
DR   EMBL; L29404; AAA83545.1; -.
DR   EMBL; X00006; CAA24909.1; -.
DR   PIR; A03509; YTECR1.
DR   InterPro; IPR001958; TCR_TetA.
DR   InterPro; IPR003662; sub_trnsportr.
DR   PRINTS; PR01035; TCRTETA.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW   Antibiotic resistance; Transmembrane; Inner membrane; Transport;
KW   Symport; Transposable element; Plasmid.
FT   DOMAIN        1      7       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM      8     27       PROBABLE.
FT   DOMAIN       28     45       PERIPLASMIC (PROBABLE).
FT   TRANSMEM     46     66       PROBABLE.
FT   DOMAIN       67     79       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM     80    100       PROBABLE.
FT   DOMAIN      101    103       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    104    124       PROBABLE.
FT   DOMAIN      125    138       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    139    159       PROBABLE.
FT   DOMAIN      160    160       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    161    181       PROBABLE.
FT   DOMAIN      182    210       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    211    231       PROBABLE.
FT   DOMAIN      232    246       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    247    267       PROBABLE.
FT   DOMAIN      268    277       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    278    298       PROBABLE.
FT   DOMAIN      299    299       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    300    320       PROBABLE.
FT   DOMAIN      321    339       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    340    360       PROBABLE.
FT   DOMAIN      361    364       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    365    385       PROBABLE.
FT   DOMAIN      386    399       CYTOPLASMIC (PROBABLE).
FT   CONFLICT      5      5       R -> I (IN REF. 3).
FT   CONFLICT     55     55       M -> V (IN REF. 3).
FT   CONFLICT     75     75       V -> I (IN REF. 3).
FT   CONFLICT     84     84       A -> T (IN REF. 3).
FT   CONFLICT    201    203       ASF -> SFV (IN REF. 3).
SQ   SEQUENCE   399 AA;  42240 MW;  298427E6B5478374 CRC64;


  Query Match             4.4%;  Score 125.5;  DB 1;  Length 399;
  Best Local Similarity   22.3%;  Pred. No. 0.053;
  Matches  104;  Conservative   57;  Mismatches  136;  Indels  169;  Gaps   27

Qy      19 LLVNLLTFGLEVCLAAGITYVPPL---LLEVGVEEKFMT----MVLGIGPVLGLVCVPLL 71
           |:| | |  |:   | ||  : |:   ||   |    :|    ::| :  ::   | |:|
Db       7 LIVILSTVALD---AVGIGLIMPVLPGLLRDLVHSNDVTAHYGILLALYALMQFACAPVL 63

Qy      72 GSASDHWRGRYGRR------------------RPFIWALSLGILLSLFLIPRAGWLAGL 112
           |: ||    |:|||                    ||:| | :| :::
Db      64 GALSD----RFGRRPVLLVSLAGAAVDYAIMATAPFLWVLYIGRIVA------------ 106

Qy     113 LCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGG 172
                       |:    | |      | ::|:    :  |     : ||
Db     107 ----------------GITGATGAVA----GAYIADITDGDERARH----FGFM---SA 138

Qy     173 CLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGL 232
           | |: : |          | ||       ||:
Db     139 CFGFGMVA--------GPVLG-------GLM---------------------------- 154

Qy     233 SAPSLSPHCCPCRARLAFRNLG-----ALLPRLHQLCCRMPRTLRRLFVAELCS--W--- 282
               |||  |  |  |   |        |||  |:      | |||  :  | |  |
Db     155 --GGFSPH-APFFAAAALNGLNFLTGCFLLPESHK---GERRPLRREALNPLASFRWARG 208

Qy     283 ---MALMTFTLFYTDFVGEGLYQGVPRA--EPGTEARRHYDE---GVRMGSLGLFLQCAI 334
             :| :    |    ||:     || |      | | |:|    |: : : |:    |
Db     209 MTVVAALMAVFFIMQLVGQ-----VPAALWVIFGEDRFHWDATTIGISLAAFGILHSLAQ 263

Qy     335 SLVFSLVMDRLVQRFGTRAVYLASVA---AFPVAAGAT--CLSHSVAVVTASAALTGFTF 389
           :::   |  || :|   ||: |  :|    : : | ||   ::  : |: ||    |
Db     264 AMITGPVAARLGER---RALMLGMIADGTGYILLAFATRGWMAFPIMVLLASG---GIGM 317

Qy     390 SALQILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGP 435
           |||  :   |:     |:|       |   |: :  : :|| : ||
Db     318 PALQAM---LSRQVDEERQ------GQLQGSLAALTSLTSIV-GP 352


RESULT   4
PUR8_STRLP
ID   PUR8_STRLP     STANDARD;      PRT;   503 AA.
AC   P42670;
DT   01-NOV-1995 (Rel. 32, Created)
```

```
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   PUROMYCIN RESISTANCE PROTEIN PUR8.
GN   PUR8.
OS   Streptomyces lipmanii (Streptomyces alboniger).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 12461;
RX   MEDLINE=94109397; PubMed=7916693;
RA   Tercero J.A., Lacalle R.A., Jimenez A.;
RT   "The pur8 gene from the pur cluster of Streptomyces alboniger encodes
RT   a highly hydrophobic polypeptide which confers resistance to
RT   puromycin.";
RL   Eur. J. Biochem. 218:963-971(1993).
CC   -!- FUNCTION: MAY BE INVOLVED IN ACTIVE PUROMYCIN EFFLUX ENERGIZED BY
CC       A PROTON-DEPENDENT ELECTROCHEMICAL GRADIENT. IN ADDITION, IT COULD
CC       BE IMPLICATED IN SECRETING N-ACETYLPUROMYCIN, THE LAST
CC       INTERMEDIATE OF THE PUROMYCIN BIOSYNTHESIS PATHWAY, TO THE
CC       ENVIRONMENT.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC       AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X76855; CAA54186.1; -.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Antibiotic resistance; Antibiotic biosynthesis; Transmembrane;
KW   Transport.
FT   DOMAIN        1     24       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     25     45       POTENTIAL.
FT   DOMAIN       46     64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     65     85       POTENTIAL.
FT   DOMAIN       86     92       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     93    113       POTENTIAL.
FT   DOMAIN      114    122       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    123    143       POTENTIAL.
FT   DOMAIN      144    152       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    153    173       POTENTIAL.
FT   DOMAIN      174    181       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    182    202       POTENTIAL.
FT   DOMAIN      203    212       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    213    233       POTENTIAL.
FT   DOMAIN      234    241       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    242    262       POTENTIAL.
FT   DOMAIN      263    280       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    281    301       POTENTIAL.
FT   DOMAIN      302    313       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    314    334       POTENTIAL.
FT   DOMAIN      335    346       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    347    367       POTENTIAL.
FT   DOMAIN      368    371       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    372    392       POTENTIAL.
FT   DOMAIN      393    422       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    423    443       POTENTIAL.
FT   DOMAIN      444    461       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    462    482       POTENTIAL.
FT   DOMAIN      483    503       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   503 AA;  51852 MW;  FA3B814DF9A3EB74 CRC64;
```

```
Query Match             4.2%;  Score 121.5;  DB 1;  Length 503;
Best Local Similarity   23.2%;  Pred. No. 0.13;
Matches  131;  Conservative   61;  Mismatches  175;  Indels  197;  Gaps   32;
```

```
Qy      26 FGLEVCLAAGITYV----------PPLLLEVGVEEKFMTMV-----LGIGPVLGLVCVPL 70
           :|| | ||| :  |          | :  ::|: :     |     |  | :|      |
Db      26 WGLVVILAAQLLVVLDGTVVNIALPSVQRDLGMSDTSRQWVITAYTLAFGGLL------L 79

Qy      71 LGSASDHWRGR----YGRRRPFIWALSLGILLSLFLIPRAGW-LAGLL---CPDPRPLEL 122
           ||       ||    :|||| |    ::|||          |: || ||    ||| | |
Db      80 LG-------GRVADAFGRRRIF----AVGIL---------GFGLASLLGGAAPDPGTLFL 119

Qy     123 ALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAID 182
           |  : ||    |   :   | ||:: || :|    :|: ||  :   |  :| |
Db     120 ARALQGV----FAAALAPAAL-ALINTLFTEPGERGKAFGVYGAVSGGGAAVGLLAG--- 171

Qy     183 WDTSALAPYLGTQEECLFGLLT--LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
                          ||||  | :  |:     ||  | ||
Db     172 -----------------GLLTEYLDWRWCLYVNAPVALLALLG---------------- 197

Qy     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAEL--CSWMALMTFTLFYTDFVGE 298
              ||          ||||      |  | :|      |  |  :  : :
Db     198 ---CR----------LLPRDR----RTGRAVRLDLPGTLLGCGGLVAIVYAF-------- 232

Qy     299 GLYQGVPRAEPGTEARRHYDEG--VRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYL 356
                     ||   : :   ||:  ||: :  | :||   | | |:
Db     233 -----------AEAESGWGDPLVVRLLVLGVLMLVAFALVERRVQDPLLP--------- 271

Qy     357 ASVAAFPVAAGATCLSHSVAVVTASAALTG----FTFSALQILPYT--LASLYHREKQVF 410
             | |  |  |:      : |      | |     |:    || |:  |  :      |
Db     272 PGVVAHRVRGGS-----FLVVGLPQIGLFGLFLFLTYYLQGILDYSPVLTGV------AF 320

Qy     411 LPKYRGDTGGASSEDSLMTSFLPGPKP-----GAPFPNGHVGAGGSGLL----PPPPALC 461
           ||   |   |:|    :    ||  :|     ||         | |  ||    |  |
Db     321 LPLGLGIAVGSS---LIAARLLPRTRPRTLIVGALL----AAAAGMALLTRLEPDTP--- 370

Qy     462 GASACDVSVRVVVGEPTEARVVPGRGI-CLDLAILDSAFLLSQVAP------SLFMGSIV 514
                 :| :     |::: | || |: :  : :|   ::|||      :  : |
Db     371 ---------QVYLTHLLPAQILIGLGIGCMMMPAMHTA--TARVAPHEAGAAAAVVNSAQ 419

Qy     515 QLSQSVTAYMVSAAGLGLVAIYFA 538
           |:  ::   :::    |  | | |
Db     420 QVGGALGVALLNTVSTGATAAYLA 443
```

```
RESULT   5
AMPG_ECOLI
ID   AMPG_ECOLI     STANDARD;      PRT;   491 AA.
AC   P36670;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   AMPG PROTEIN.
GN   AMPG OR B0433 OR Z0536 OR ECS0487.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94049112; PubMed=8231804;
RA   Lindquist S., Weston-Hafer K., Schmidt H., Pul C., Korfmann G.,
RA   Erickson J., Sanders C., Martin H.H., Normark S.;
RT   "AmpG, a signal transducer in chromosomal beta-lactamase induction.";
RL   Mol. Microbiol. 9:703-715(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SN0301-1, SN0301-3, AND SN0301-5;
RX   MEDLINE=95291453; PubMed=7773404;
RA   Schmidt H., Korfmann G., Barth H., Martin H.H.;
RT   "The signal transducer encoded by ampG is essential for induction of
RT   chromosomal AmpC beta-lactamase in Escherichia coli by beta-lactam
RT   antibiotics and 'unspecific' inducers.";
RL   Microbiology 141:1085-1092(1995).
RN   [3]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA   Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA   Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- FUNCTION: PROBABLY ACTS AS A PERMEASE IN THE BETA-LACTAMASE
CC       INDUCTION SYSTEM AND IN PEPTIDOGLYCAN RECYCLING.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (PROBABLE).
CC   -!- SIMILARITY: TO H.INFLUENZAE HI0350 AND TO YEAST YBR220C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S67816; AAB28884.1; -.
DR   EMBL; X82158; CAA57651.1; -.
DR   EMBL; X82159; CAA57652.1; -.
DR   EMBL; X82160; CAA57653.1; -.
DR   EMBL; AE000149; AAC73536.1; -.
DR   EMBL; U82664; AAB40189.1; -.
DR   EMBL; AE005222; AAG54783.1; -.
DR   EMBL; AP002551; BAB33910.1; -.
DR   PIR; S37391; S37391.
DR   EcoGene; EG12183; ampG.
KW   Transport; Transmembrane; Inner membrane; Complete proteome.
FT   TRANSMEM     13     33       POTENTIAL.
FT   TRANSMEM     81    105       POTENTIAL.
FT   TRANSMEM    110    128       POTENTIAL.
FT   TRANSMEM    174    197       POTENTIAL.
FT   TRANSMEM    224    244       POTENTIAL.
FT   TRANSMEM    266    283       POTENTIAL.
FT   TRANSMEM    328    345       POTENTIAL.
FT   TRANSMEM    382    406       POTENTIAL.
FT   TRANSMEM    425    450       POTENTIAL.
FT   TRANSMEM    461    485       POTENTIAL.
FT   VARIANT     151    151       G -> D (IN NON-INDUCIBLE MUTANT
FT                                SN0301-1).
FT   VARIANT     268    268       G -> D (IN NON-INDUCIBLE MUTANT
FT                                SN0301-3).
FT   VARIANT     373    373       G -> D (IN NON-INDUCIBLE MUTANT
FT                                SN0301-5).
SQ   SEQUENCE   491 AA;  53245 MW;  6B0D7E2BC699CA9C CRC64;


  Query Match             4.2%;  Score 119.5;  DB 1;  Length 491;
  Best Local Similarity   21.0%;  Pred. No. 0.18;
  Matches  126;  Conservative   76;  Mismatches  193;  Indels  205;  Gaps   30

Qy       9 RLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTM------VLGIGPV 62
           |: :  :: :||:         :  |:|:    || |  |  : :||:       :|  :
Db       7 RIFQQPRSAILLI--------LGFASGL----PLALTSGTLQAWMTVENIDLKTIGFFSL 54

Qy      63 LGLVCV------PLLGSASDHWRGRYGRRRPFIWALSLGILLSL----FLIP--RAGWLA 110
           :|   |       ||:   :  :   :|||| :: |  : :|:::    || |  :  |:|
Db      55 VGQAYVFKFLWSPLMDRYTPPF---FGRRRGWLLATQILLLVAIAAMGFLEPGTQLRWMA 111

Qy     111 GLLCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDH-CRQAYSVYAF--- 166
           |                :: ||        :|  :|:    :     | ||  :
Db     112 ALAV---------------VIAFCSASQDIVFDAWKTDVLPAEERGAGAAISVLGYRLG 155

Qy     167 MISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPT 226
           |:  ||   :|  |  |        || |   :: |:  :  : |: ||||  |     ||
Db     156 MLVSGGLALWL--ADKW--------LGWQG--MYWLMAALLIPCIIATLLAPE-----PT 198

Qy     227 EPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAEL------- 279
           :                              :|        :|:|| :  || |
Db     199 D----------------------------TIP--------VPKTLEQAVVAPLRDFFGRN 222

Qy     280 CSWMALMTFTL------FYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCA 333
             :|: |:   |       |         | :||      | :|    : ||   :|||
Db     223 NAWLILLLIVLYKLGDAFAMSLTTTFLIRGV-----GFDAG---EVGVVNKTLGLLATIV 274

Qy     334 ISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQ 393
           :|   ::| ||       ||: :  :                              | | : :
Db     275 GALYGGILMQRLSL---FRALLIFGI-----------------------LQGASNAGYW 307

Qy     394 ILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGL 453
           :|  |  ||      ||      |  | ::    |||                |    |    |
Db     308 LLSITDKHLYSMGAAVFFENLCGGMGTSAFVALLMT----------LCNKSFSATQFAL 356

Qy     454 LPPPPALCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVA---PSLFM 510
           |         ||      || || |     |   |          | | |  ||   | | :
Db     357 L---------SALSAVGRVYVG-PVAGWFVEAHGW--------STFYLFSVAAAVPGLIL 398

Qy     511 GSIVQLSQ----------SVTAYMVS--------AAGLGLVAIYFATQVVFDKSDLAKYS 552
             : : :            | |||           |||: |:|::    :  |  ||  :|
Db     399 LLVCRQTLEYTRVNDNFISRTAYPAGYAFAMWTLAAGVSLLAVWLLL-LTMDALDLTHFS 457


RESULT   6
SMVA_SALTY
ID   SMVA_SALTY     STANDARD;      PRT;   496 AA.
AC   P37594;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   METHYL VIOLOGEN RESISTANCE PROTEIN SMVA.
GN   SMVA.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SL1303;
RX   MEDLINE=95011654; PubMed=7926834;
RA   Hongo E., Morimyo M., Mita K., Machida I., Hama-Inaba H., Tsuji H.,
RA   Ichimura S., Noda Y.;
```

```
RT   "The methyl viologen-resistance-encoding gene smvA of Salmonella
RT   typhimurium.";
RL   Gene 148:173-174(1994).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC       AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D26057; BAA05055.1; -.
DR   StyGene; SG10384; smvA.
KW   Transmembrane; Inner membrane.
FT   TRANSMEM      5     25       POTENTIAL.
FT   TRANSMEM     44     64       POTENTIAL.
FT   TRANSMEM     73     93       POTENTIAL.
FT   TRANSMEM     96    116       POTENTIAL.
FT   TRANSMEM    135    155       POTENTIAL.
FT   TRANSMEM    158    178       POTENTIAL.
FT   TRANSMEM    192    212       POTENTIAL.
FT   TRANSMEM    220    240       POTENTIAL.
FT   TRANSMEM    260    280       POTENTIAL.
FT   TRANSMEM    299    319       POTENTIAL.
FT   TRANSMEM    327    347       POTENTIAL.
FT   TRANSMEM    357    377       POTENTIAL.
FT   TRANSMEM    391    411       POTENTIAL.
FT   TRANSMEM    470    490       POTENTIAL.
SQ   SEQUENCE   496 AA;  52521 MW;  0F1B23C8FD27BCCD CRC64;
```

```
  Query Match             4.1%;  Score 118.5;  DB 1;  Length 496;
  Best Local Similarity   22.5%;  Pred. No. 0.21;
  Matches  102;  Conservative   44;  Mismatches  143;  Indels  165;  Gaps   22;

Qy      17 QLLLVNLLTFGL----------------EVCLAAGITYVPPLLLEVGV-----EEKFMT 54
           :||::    |||                 | || |   : |  |  |:     |||
Db      72 RLLMLGGTLFGLASLAAAFSHTASWLIATRVLLAIGAAMIVPATL-AGIRATFCEEKHRN 130

Qy      55 MVLGIGPVL---GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILL------SLFLI-- 103
           | ||:   :   |   |||:|                    ||||      |:|||
Db     131 MALGVWAAVGSGGRAFVPLIG--------------------GILLEHFYWGSVFLINV 168

Qy     104 PRAGWLAGLLCP-DPR-------PLEL---ALLILGVGLLDFCGQ----------VCFT 141
           |    : ||    |||       || |    :||: : || :  :          : :|
Db     169 PIVLVVMGLTARYDPRQAGRRDQPLNLGHVVMLIIAILLLVYSAKTALKGHLSLWVISYT 228

Qy     142 PLE-ALLSDLFRDPD---------------------------------------------- 155
           |  |||  ||
Db     229 LLTGALLLGLFIRTQLATSRPMIDMRLFTHRIILSGVVMAMTAMITLVGFELLMAQELQF 288

Qy     156 -HCRQAYSVYAFMI----------SLGGCL------------GYLLPAIDWDTSALAPYL 192
           |    |    ||:           : | |            |  | |: :   |:  :
Db     289 VHGLSPYEAGVFMLPVMVASGFSGPIAGVLVSRLGLRLVATGGMALSALSFYGLAMTDF- 347

Qy     193 GTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAE-GLSAPSLSPHCCPCRARLAFR 251
           ||:   :||: |:  :  ||: |:|  :|:    |||   :| ::       | |
Db     348 STQQWQAWGLMALLGFS--AASALLASTSAIMAAAPAEKAAAAGAIETMAYELGAGLGIA 405

Qy     252 NLGALLPRLHQLCCRMPRTLRRLFVAELCSWMA-LMTFTLFYTDFVGEGLYQGVPRAEPG 310
           | || |      |:|  |    :|   | |   :    |     |:| |
Db     406 IFGLLLSRSFSASIRLPAGLEAQEIARASSSMGEAVQLANSYPPTQGQGKY--------- 456

Qy     311 TEARRH---YDEGVRMGSLG-LFLQCAISLVFSL 340
           | ||   :   | : | | : |  |: : |||
Db     457 LTAARHAFIWSHSVALSSAGSMLLLLAVGMWFSL 490
```

```
RESULT   7
YHJE_ECOLI
ID   YHJE_ECOLI     STANDARD;      PRT;   440 AA.
AC   P37643;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YHJE.
GN   YHJE OR B3523.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94316500; PubMed=8041620;
RA   Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT   "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT   region from 76.0 to 81.5 minutes.";
RL   Nucleic Acids Res. 22:2576-2586(1994).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. STRONG, TO
CC       H.INFLUENZAE HI0281 AND HI0418.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00039; AAB18499.1; -.
DR   EMBL; AE000429; AAC76548.1; -.
DR   EcoGene; EG12249; yhjE.
DR   InterPro; IPR003662; sub_trnsportr.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
KW   Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW   Complete proteome.
FT   TRANSMEM     35     55       POTENTIAL.
FT   TRANSMEM     67     87       POTENTIAL.
FT   TRANSMEM    109    129       POTENTIAL.
FT   TRANSMEM    130    150       POTENTIAL.
FT   TRANSMEM    168    188       POTENTIAL.
FT   TRANSMEM    193    213       POTENTIAL.
FT   TRANSMEM    249    269       POTENTIAL.
FT   TRANSMEM    290    310       POTENTIAL.
FT   TRANSMEM    321    341       POTENTIAL.
FT   TRANSMEM    346    366       POTENTIAL.
FT   TRANSMEM    385    405       POTENTIAL.
FT   TRANSMEM    411    431       POTENTIAL.
SQ   SEQUENCE   440 AA;  47208 MW;  3A23302A9DCBE5BE CRC64;
```

```
  Query Match             4.1%;  Score 117;  DB 1;  Length 440;
  Best Local Similarity   28.4%;  Pred. No. 0.24;
  Matches   48;  Conservative   29;  Mismatches   54;  Indels   38;  Gaps   10;

Qy      53 MTMVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLI-PRAGWLAG 111
           |  |:| |     | ||: |  :|     :|||:  :   :| || :||   |  |
Db     295 MMAVIGFG-----VMVPVAGLLAD----AFGRRKSMVIITTLIILFALFAFNPLLG---- 341

Qy     112 LLCPDPRP-LELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISL 170
                 | |  | |:||: |:     : | |: ||| :||  |   |  |:  :|  ::
Db     342 ----SGNPILVFAFLLLGLSLMG----LTFGPMGALLPELF--PTEVR--YTGASFSYNV 389

Qy     171 GGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLTL-IFLTCVAATLLVA 218
               ||          :::|||:    :  :||  : ::|  :|   |:|
Db     390 ASILG----------ASVAPYIAAWLQTNYGLGAVGLYLAAMAGLTLIA 428
```

```
RESULT   8
Y051_MYCTU
ID   Y051_MYCTU     STANDARD;      PRT;   640 AA.
AC   Q10880; O53628;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL 65.6 KDA PROTEIN RV0083.
GN   RV0083 OR MT0090 OR MTCY251.01 OR MTV030.27.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
CC       OF CHLOROPLASTS OR MITOCHONDRIA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021428; CAA16264.1; -.
DR   EMBL; AE006920; AAK44315.1; -.
DR   TIGR; MT0090; -.
DR   TubercuList; Rv0083; -.
DR   InterPro; IPR001750; Oxidored_q1.
DR   Pfam; PF00361; oxidored_q1; 1.
KW   Hypothetical protein; Oxidoreductase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     52     72       POTENTIAL.
FT   TRANSMEM     90    110       POTENTIAL.
FT   TRANSMEM    136    156       POTENTIAL.
FT   TRANSMEM    179    199       POTENTIAL.
FT   TRANSMEM    208    228       POTENTIAL.
FT   TRANSMEM    241    261       POTENTIAL.
FT   TRANSMEM    277    297       POTENTIAL.
FT   TRANSMEM    298    318       POTENTIAL.
FT   TRANSMEM    352    372       POTENTIAL.
FT   TRANSMEM    391    411       POTENTIAL.
FT   TRANSMEM    446    466       POTENTIAL.
FT   TRANSMEM    497    517       POTENTIAL.
FT   TRANSMEM    619    639       POTENTIAL.
SQ   SEQUENCE   640 AA;  65627 MW;  6254C289DBD108A8 CRC64;


  Query Match             3.9%;  Score 111.5;  DB 1;  Length 640;
  Best Local Similarity   22.3%;  Pred. No. 0.86;
  Matches  129;  Conservative   73;  Mismatches  222;  Indels  155;  Gaps   29

Qy      16 AQLLLV----NLLTFGL--EVCLAAGITYV-----PPLLLEVGVEEKFMTMVLGIGPVLG 64
           | :|||    :: || |  |:   | :  |      | :   |:    || :  |  ::|
Db      93 AAMLLVPAAGSVTTFLLAWELMAIASLILVLSEHARPQVRSAGLWYAVMTQLGFIAILVG 152

Qy      65 LVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLF-----------LIPRAGWLAGL 112
           || :   |  || : |        : |:  |:  ::|            |:|   ||
Db     153 LVVLAAAG-GSDRFAG--------LGAVCDGVRAAVFMLTLVGFGSKAGLVPLHAWLPRA 203

Qy     113 LCPDPRP----LELALLILGV-GLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFM 167
             | |    :  |:: ||: |:: |  |:           |   |     |      :
Db     204 HPEAPSPVSALMSAAMVNLGIYGIVRFDLQL----------LGPGPRWWGLA------L 246

Qy     168 ISLGG---CLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALG 224
           :::||     | |  ::  |   |  |  |:     ||:||     : |  | |: | |
Db     247 LAVGGTSALYGVLQASVAADLKRLLAYSTTEN---MGLITL----ALGAATLFADTGAYG 299

Qy     225 PTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPR--------LHQLCCRMPRTLRRLFV 276
           |   |   :   :  |       |||   |::|           |  |  ||| |
Db     300 PASIAAAAAMLHMIAHA--AFKSLAFMAAGSVLAATGLRDLDLLGGLARRMPAT------ 351

Qy     277 AELCSWMALMTFTLFY-------------TDFVGEG-LYQGVPRAEPGTEARRHYDEGVR 322
                   |:|:                || |  | | :  | || :        :
Db     352 ------------TVFFGVAALGACGLPLGAGFVSEWLLVQSLIHAAPGHDPIVALTTPLA 399

Qy     323 MGSLGLFLQCAISLVFS------LVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVA 376
           :| : |    ::: :        |  |  |  |    :  |  :|||| ||  :||
Db     400 VGVVALATGLSVAAMTKAFGIGFLARPRSTQAEAAREAPASMRAGMAIAAGA-CLVLAVA 458

Qy     377 --VVTASAALTGFTFSALQILPYT-LASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLP 433
             :|          |  | | : :| | ::      | ||   |              |  |
Db     459 PLLVAPMVRRAAATLPAAQAVKFTGLGAV------VRLPAMSG-------------SIAP 499

Qy     434 GPKPGAPFPNGHVGAGGSGLLP-------PPPA-----LCGASACDVSVRVVVGEPTEAR 481
           |    |        |    :|        | ||      |||:  |::||:     : |
Db     500 GVIAAAVL----AAALAVAVLARWRFRRRPAPARLPLWACGAA--DLTVRMQYTATSFAE 553

Qy     482 VVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSV 520
             | : :  |:   |:   ::  | | :|   :    :|
Db     554 --PLQRVFGDVLRPDTDIEVTHTAESRYMAERITYRTAV 590


RESULT   9
PHDK_NOCSK
ID   PHDK_NOCSK     STANDARD;      PRT;   473 AA.
AC   O24723;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PROBABLE 1-HYDROXY-2-NAPHTHOATE TRANSPORTER.
OS   Nocardioides sp. (strain KP7).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Propionibacterineae; Nocardioidaceae; Nocardioides.
OX   NCBI_TaxID=35761;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97474276; PubMed=9335300;
RA   Iwabuchi T., Harayama S.;
RT   "Biochemical and genetic characterization of 2-carboxybenzaldehyde
RT   dehydrogenase, an enzyme involved in phenanthrene degradation by
RT   Nocardioides sp. strain KP7.";
RL   J. Bacteriol. 179:6488-6494(1997).
CC   -!- FUNCTION: PROBABLY INVOLVED IN THE TRANSPORT OF 1-HYDROXY-2-
CC       NAPHTHOATE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
```

```
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB000735; BAA23264.1; -.
DR   InterPro; IPR003662; sub_trnsportr.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
KW   Transport; Transmembrane; Inner membrane.
FT   TRANSMEM     26     46       POTENTIAL.
FT   TRANSMEM     60     80       POTENTIAL.
FT   TRANSMEM     92    112       POTENTIAL.
FT   TRANSMEM    122    142       POTENTIAL.
FT   TRANSMEM    154    174       POTENTIAL.
FT   TRANSMEM    181    201       POTENTIAL.
FT   TRANSMEM    264    284       POTENTIAL.
FT   TRANSMEM    305    325       POTENTIAL.
FT   TRANSMEM    331    351       POTENTIAL.
FT   TRANSMEM    355    375       POTENTIAL.
FT   TRANSMEM    416    436       POTENTIAL.
SQ   SEQUENCE   473 AA;  49109 MW;  D6D765D376260D8A CRC64;


  Query Match             3.9%;  Score 111;  DB 1;  Length 473;
  Best Local Similarity   24.3%;  Pred. No. 0.68;
  Matches  114;  Conservative   66;  Mismatches  180;  Indels  110;  Gaps   27;

Qy      12 RHRKAQLLLVNLLTF--GLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVP 69
           | ::  :|:| |:    |::: ||: :   ||:: : ||    :|:|: :| |   :
Db      22 RRQRIIVLMVFLIMVADGMDITLASHL--FPPVIRDWGVPVSAVTLVVSLGVVAMAIGAL 79

Qy      70 LLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAG-WLAGLLCPDPRPLELALLILG 128
           : |  :| | || |          :| :  || :  ||  | |    |        :|
Db      80 VSGPVADRW-GRKG-------VTVVGFV--LFCLATAGLGLTG----DIHSFAALRIISC 125

Qy     129 VGLLDFCGQVCFTPLEALLSDLFRDPDHCR-QAYSVYAFMISLGGCLGYLLPAIDWDTSA 187
            ||    | |   |  :::|    |   | |  |:    : :|  :|  | |     |
Db     126 FGL----GAVMPVAL-TIVADWM--PKARRAQMVSIAFAGVGVGSIIGAYLAA------A 172

Qy     188 LAPYLGTQEECLF-GLLTLIFLTCVAATLLVAEEA---------------ALGPTEPAEG 231
           : | || |   |  ||  || |    |  || | |                ||    |
Db     173 VIPTLGWQVMVLIAGLAPLIILPFFVA--LVPEPAIISVRRGIPEARIRSALALVAPDRD 230

Qy     232 LSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLF 291
           ::   |:      || |    |||   |            : ||   ||  :  :|  ::
Db     231 IAGVDLT------RAGL---TLGAGEVR----------AKALFAEILCRPLLGVTLLIW 270

Qy     292 YTDFVGEG----LYQGVP----RAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMD 343
              || :|    : | :|      ||       : |: :   |       :|:  | :
Db     271 GVFFVVQGSGLLVLQYMPMLLQAPAPGLST---VESGLIVAMYGWG-----ALIGQLTIA 322

Qy     344 RLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTF-------SALQI-L 395
            :::||  | : ||:   : |          | ::  :|  ||| |         |: : |
Db     323 FILKRF-DRFIALAAFIFWSV----------VGLLIVAAFGTGFGFFGYFTLLFAIGLSL 371

Qy     396 PYTLASLYHREKQVFLPKYR----GDTGGASSEDSLMTSFLPGPKPGAPF 441
           | | |::       :  ::|    |  | |    :|    | |   || |
Db     372 PATAAAMQSVTTLAYEEEFRATGMGSAGFAGRLGTLTYGALGGTLIGAGF 421


RESULT  10
CA1A_HUMAN
ID   CA1A_HUMAN     STANDARD;      PRT;   680 AA.
AC   Q03692;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(X) CHAIN PRECURSOR.
GN   COL10A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92109659; PubMed=1764025;
RA   Thomas J.T., Cresswell C.J., Rash B., Nicolai H., Jones T.,
RA   Solomon E., Grant M.E., Boot-Handford R.P.;
RT   "The human collagen X gene. Complete primary translated sequence and
RT   chromosomal localization.";
RL   Biochem. J. 280:617-623(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93012005; PubMed=1397333;
RA   Reichenberger E., Beier F., Luvalle P., Olsen B.R., von der Mark K.,
RA   Bertling W.M.;
RT   "Genomic organization and full-length cDNA sequence of human collagen
RT   X.";
RL   FEBS Lett. 311:305-310(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Beier F., Lammi M.B., von der Mark K.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 52-680 FROM N.A.
RX   MEDLINE=92267014; PubMed=1587271;
RA   Apte S.S., Seldin M.F., Hayashi M., Olsen B.R.;
RT   "Cloning of the human and mouse type X collagen genes and mapping of
RT   the mouse type X collagen gene to chromosome 10.";
RL   Eur. J. Biochem. 206:217-224(1992).
RN   [5]
RP   SEQUENCE OF 561-666 FROM N.A.
RX   MEDLINE=91243838; PubMed=2037056;
RA   Apte S., Mattei M.-G., Olsen B.R.;
RT   "Cloning of human alpha 1(X) collagen DNA and localization of the
RT   COL10A1 gene to the q21-q22 region of human chromosome 6.";
RL   FEBS Lett. 282:393-396(1991).
RN   [6]
RP   SEQUENCE OF 547-655 FROM N.A.
RX   MEDLINE=92077285; PubMed=1743401;
RA   Reichenberger E., Aigner T., von der Mark K., Stoeb H., Bertling W.;
RT   "In situ hybridization studies on the expression of type X collagen
RT   in fetal human cartilage.";
RL   Dev. Biol. 148:562-572(1991).
RN   [7]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255959; PubMed=9101290;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   Hum. Mutat. 9:300-315(1997).
RN   [8]
RP   VARIANTS SMCD ASP-598 AND PRO-614.
RX   MEDLINE=94136476; PubMed=8304336;
RA   Wallis G.A., Rash B., Sweetman W.A., Thomas J.T., Super M., Evans G.,
RA   Grant M.E., Boot-Handford R.P.;
RT   "Amino acid substitutions of conserved residues in the
RT   carboxyl-terminal domain of the alpha 1(X) chain of type X collagen
RT   occur in two unrelated families with metaphyseal chondrodysplasia
RT   type Schmid.";
RL   Am. J. Hum. Genet. 54:169-178(1994).
RN   [9]
RP   VARIANT SMCD ARG-591.
RX   MEDLINE=94272470; PubMed=8004099;
RA   McIntosh I., Abbott M.H., Warman M.L., Olsen B.R., Francomano C.A.;
RT   "Additional mutations of type X collagen confirm COL10A1 as the
RT   Schmid metaphyseal chondrodysplasia locus.";
```

```
RL   Hum. Mol. Genet. 3:303-307(1994).
RN   [10]
RP   VARIANT SMCD VAL-618.
RX   MEDLINE=95181449; PubMed=7876225;
RA   Chan D., Cole W.G., Rogers J.G., Bateman J.F.;
RT   "Type X collagen multimer assembly in vitro is prevented by a Gly618
RT   to Val mutation in the alpha 1(X) NC1 domain resulting in Schmid
RT   metaphyseal chondrodysplasia.";
RL   J. Biol. Chem. 270:4558-4562(1995).
RN   [11]
RP   VARIANTS SMCD ARG-545; GLU-595; HIS-597; LYS-617; ARG-644 AND GLY-648.
RX   MEDLINE=95331767; PubMed=7607655;
RA   Bonaventure J., Chaminade F., Maroteaux P.;
RT   "Mutations in three subdomains of the carboxy-terminal region of
RT   collagen type X account for most of the Schmid metaphyseal
RT   dysplasias.";
RL   Hum. Genet. 96:58-64(1995).
RN   [12]
RP   VARIANT SMCD PRO-600.
RX   MEDLINE=96375754; PubMed=8782043;
RA   Wallis G.A., Rash B., Sykes B., Bonaventure J., Maroteaux P.,
RA   Zabel B., Wynne-Davies R., Grant M.E., Boot-Handford R.P.;
RT   "Mutations within the gene encoding the alpha 1 (X) chain of type X
RT   collagen (COL10A1) cause metaphyseal chondrodysplasia type Schmid but
RT   not several other forms of metaphyseal chondrodysplasia.";
RL   J. Med. Genet. 33:450-457(1996).
RN   [13]
RP   VARIANTS SMCD GLU-18 AND ARG-18.
RX   MEDLINE=97220591; PubMed=9067753;
RA   Ikegawa S., Nakamura K., Nagano A., Haga N., Nakamura Y.;
RT   "Mutations in the N-terminal globular domain of the type X collagen
RT   gene (COL10A1) in patients with schmid metaphyseal
RT   chondrodysplasia.";
RL   Hum. Mutat. 9:131-135(1997).
RN   [14]
RP   VARIANTS SMD GLU-595.
RX   MEDLINE=99057503; PubMed=9837818;
RA   Ikegawa S., Nishimura G., Nagai T., Hasegawa T., Ohashi H.,
RA   Nakamura Y.;
RT   "Mutation of the type X collagen gene 'COL10A1' causes
RT   spondylometaphyseal dysplasia.";
RL   Am. J. Hum. Genet. 63:1659-1662(1998).
RN   [15]
RP   VARIANT SMCD CYS-597.
RX   MEDLINE=99069781; PubMed=9852679;
RA   Sawai H., Ida A., Nakata Y., Koyama K.;
RT   "Novel missense mutation resulting in the substitution of tyrosine by
RT   cysteine at codon 597 of the type X collagen gene associated with
RT   Schmid metaphyseal chondrodysplasia.";
RL   J. Hum. Genet. 43:259-261(1998).
CC   -!- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTHROPHIC
CC       CHONDROTOCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
CC       MINERALIZATION ZONES OF HYALINE CARTILAGE.
CC   -!- SUBUNIT: HOMOTRIMER.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SCHMID TYPE
CC       METAPHYSEAL CHONDRODYSPLASIA (SMCD), A DOMINANTLY INHERITED
CC       DISORDER OF THE OSSEOUS SKELETON. THE CARDINAL FEATURES OF THE
CC       PHENOTYPE ARE MILD SHORT STATURE, COXA VARA AND A WADDLING GAIT.
CC       RADIOGRAPHY USUALLY SHOWS SCLEROSIS OF THE RIBS, FLARING OF THE
CC       METAPHYSES, AND A WIDE IRREGULAR GROWTH PLATE, ESPECIALLY OF THE
CC       KNEES.
CC   -!- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SPONDYLOMETAPHYSEAL
CC       DYSPLASIA (SMD). SMD COMPRISES A HETEROGENEOUS GROUP OF HERITABLE
CC       SKELETAL DYSPLASIAS CHARACTERIZED BY MODIFICATIONS OF THE
CC       VERTEBRAL BODIES OF THE SPINE AND METAPHYSES OF THE TUBULAR BONES.
CC   -!- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
CC   -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X60382; CAA42933.1; -.
DR   EMBL; X65120; CAA46236.1; -.
DR   EMBL; X98568; CAA67178.1; -.
DR   EMBL; S68531; AAC60615.1; -.
DR   EMBL; X58879; CAA41686.1; -.
DR   EMBL; M74050; AAA61221.1; -.
DR   EMBL; X72579; CAA51170.1; -.
DR   EMBL; X72580; CAA51170.1; JOINED.
DR   PIR; S15826; S15826.
DR   PIR; S30086; S30086.
DR   PIR; A43901; A43901.
DR   PIR; S18249; S18249.
DR   PIR; S21856; S21856.
DR   PIR; S26396; S26396.
DR   MIM; 120110; -.
DR   MIM; 156500; -.
DR   MIM; 184250; -.
DR   InterPro; IPR001073; C1q.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF00386; C1q; 1.
DR   Pfam; PF01391; Collagen; 7.
DR   PRINTS; PR00007; COMPLEMNTC1Q.
DR   SMART; SM00110; C1Q; 1.
DR   PROSITE; PS01113; C1Q; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cartilage; Collagen; Signal; Disease mutation; Polymorphism.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    680       COLLAGEN ALPHA 1(X) CHAIN.
FT   DOMAIN       19     56       NONHELICAL REGION (NC2).
FT   DOMAIN       57    519       TRIPLE-HELICAL REGION.
FT   DOMAIN      520    680       NONHELICAL REGION (NC1).
FT   DOMAIN      545    680       C1Q.
FT   VARIANT      18     18       G -> E (IN SMCD).
FT                                /FTId=VAR_001838.
FT   VARIANT      18     18       G -> R (IN SMCD).
FT                                /FTId=VAR_001839.
FT   VARIANT     545    545       G -> R.
FT                                /FTId=VAR_001840.
FT   VARIANT     591    591       C -> R (IN SMCD).
FT                                /FTId=VAR_001841.
FT   VARIANT     595    595       G -> E (IN SMCD AND SMD).
FT                                /FTId=VAR_001842.
FT   VARIANT     597    597       Y -> C (IN SMCD).
FT                                /FTId=VAR_008039.
FT   VARIANT     597    597       Y -> H (IN SMCD).
FT                                /FTId=VAR_001843.

  Query Match             3.9%;  Score 110.5;  DB 1;  Length 680;
  Best Local Similarity   30.7%;  Pred. No. 1.1;
  Matches   43;  Conservative   11;  Mismatches   51;  Indels   35;  Gaps    7;

Qy     412 PKYRGDTG--GASSEDSLMTSFLPGP--KPGAPFPNGHVGAGGSGLLPPPPALCGASACD 467
           |  :|| |  |      : |  | ||   || | | || |  |   || ||
Db     463 PGSKGDPGSPGPPGPAGIATKGLNGPTGPPGPPGPRGHSGEPG---LPGPP--------- 510

Qy     468 VSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSA 527
               | |      ||:      |::   |: :   |||    :|  :| ||   |||
Db     511 -------GPPGP----PGQ------AVMPEGFIKAGQRPSLSGTPLVSANQGVTGMPVSA 553

Qy     528 AGLGLVAIY--FATQVVFDK 545
           : |   |    | : |||
Db     554 FTVILSKAYPAIGTPIPFDK 573


RESULT  11
GALT_HUMAN
ID   GALT_HUMAN     STANDARD;      PRT;   368 AA.
AC   O60755;
```

```
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   GALANIN RECEPTOR TYPE 3 (GAL3-R) (GALR3).
GN   GALR3 OR GALNR3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98389766; PubMed=9722565;
RA   Smith K.E., Walker M.W., Artymyshyn R., Bard J., Borowsky B.,
RA   Tamm J.A., Yao W.-J., Vaysse P.J.-J., Branchek T.A., Gerald C.,
RA   Jones K.A.;
RT   "Cloned human and rat galanin GALR3 receptors: pharmacology and
RT   activation of G-protein inwardly rectifying K+ channels.";
RL   J. Biol. Chem. 273:23321-23326(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RA   Bennett M.M., Lescoe M.K., Gallipoli P.Z., Ramabhadran T.V.;
RT   "Homologue of the human galanin 2 receptor gene isolated from a human
RT   uterus cDNA library.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99048961; PubMed=9832121;
RA   Kolakowski L.F. Jr., O'Neill G.P., Howard A.D., Broussard S.R.,
RA   Sullivan K.A., Feighner S.D., Sawzdargo M., Nguyen T., Kargman S.,
RA   Shiao L.-L., Hreniuk D.L., Tan C.P., Evans J., Abramovitz M.,
RA   Chateauneuf A., Coulombe N., Ng G., Johnson M.P., Tharian A.,
RA   Khoshbouei H., George S.R., Smith R.G., O'Dowd B.F.;
RT   "Molecular characterization and expression of cloned human galanin
RT   receptors GALR2 and GALR3.";
RL   J. Neurochem. 71:2239-2251(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99127120; PubMed=9928159;
RA   Iismaa T.P., Fathi Z., Hort Y.J., Iben L.G., Dutton J.L., Baker E.,
RA   Sutherland G.R., Shine J.;
RT   "Structural organization and chromosomal localization of three human
RT   galanin receptor genes.";
RL   Ann. N.Y. Acad. Sci. 863:56-63(1998).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20057165; PubMed=10591208;
RA   Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA   Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.,
RA   Bagguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
RA   Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
RA   Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
RA   Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
RA   Coville G.J., Cox A.V., Davis J., Dawson E., Dhami P.D., Dockree C.,
RA   Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
RA   Fleming K French L., Garner A.A., Gilbert J.G.R., Goward M.E.,
RA   Grafham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,
RA   Heathcott R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,
RA   Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,
RA   Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,
RA   Matthews L., Mccann O.T., Mcclay J., Mclaren S., Mcmurray A.A.,
RA   Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,
RA   Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,
RA   Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
RA   Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
RA   Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,
RA   Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
RA   Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
RA   Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
RA   Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
RA   Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,
RA   Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,
RA   Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E.,
RA   Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,
RA   Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,
RA   Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N.,
RA   Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,
RA   Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,
RA   Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,
RA   Rohlfing T., Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K.,
RA   Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R.,
RA   Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA   Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA   Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA   Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
RA   O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA   Khan A.S., Lane L., Tilahun Y., Wright H.;
RT   "The DNA sequence of human chromosome 22.";
RL   Nature 402:489-495(1999).
CC   -!- FUNCTION: RECEPTOR FOR THE HORMONE GALANIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF073799; AAC35944.1; -.
DR   EMBL; AF067733; AAC18860.1; -.
DR   EMBL; AF129514; AAD47348.2; -.
DR   EMBL; AF129513; AAD47348.2; JOINED.
DR   EMBL; Z97630; CAB42831.1; -.
DR   MIM; 603692; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   InterPro; IPR003908; Galnin3_rcptor .
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PRINTS; PR00663; GALANINR.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Multigene family; Lipoprotein; Palmitate.
FT   DOMAIN       1     20       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    21     41       1 (POTENTIAL).
FT   DOMAIN      42     57       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    58     78       2 (POTENTIAL).
FT   DOMAIN      79     96       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    97    118       3 (POTENTIAL).
FT   DOMAIN     119    138       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   139    159       4 (POTENTIAL).
FT   DOMAIN     160    184       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   185    205       5 (POTENTIAL).
FT   DOMAIN     206    236       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   237    257       6 (POTENTIAL).
FT   DOMAIN     258    259       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   260    280       7 (POTENTIAL).
FT   DOMAIN     281    368       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     6      6       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    95    172       BY SIMILARITY.
FT   LIPID      308    308       PALMITATE (BY SIMILARITY).
FT   DOMAIN     218    225       POLY-ALA.
FT   DOMAIN     310    318       POLY-ARG.
SQ   SEQUENCE   368 AA;  39573 MW;  E8F46B66422C3150 CRC64;


  Query Match             3.8%;  Score 110;  DB 1;  Length 368;
  Best Local Similarity   21.9%;  Pred. No. 0.62;
  Matches   94;  Conservative   35;  Mismatches  145;  Indels  156;  Gaps   17;

Qy      87 PFIWALSLGILLSLFLIPRAG---WLAGLLCPDPRPLE-----LALLILGVGLLDFCGQV 138
           | ::||       :||:   |    || || | |   :       | || : : | |  :
Db      21 PVVFAL-------IFLLGTVGNGLVLAVLLQPGPSAWQEPGSTTDLFILNLAVADLCFIL 73
```

```
Qy    139 CFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEEC 198
Db     74 CCVPFQATI---------------------------YTLDA--W--------------- 88

Qy    199 LFGLLT------LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRN 252
Db     89 LFGALVCKAVHLLIYLTMYASSFT--------------LAAVSVDRY-------LAVR- 125

Qy    253 LGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTE 312
Db    126 --------HPLRSRALRTPRNARAAVGLVWLLAALFSAPYLSYYGTVRYGALELCVPAWE 177

Qy    313 ARRHYDEGVRMGSLGLFLQCA-ISLVFSLVMDRLVQRFG----------TRAVYLASVAA 361
Db    178 DARRRALDVATFAAGYLLPVAVVSLAYGRTLRFLWAAVGPAGAAAAEARRRATGRAGRAM 237

Qy    362 FPVAA-----------------GATCLSHSVAVVTASAALTGFTFSALQILPYTLASLY 403
Db    238 LAVAALYALCWGPHHALILCFWYGRFAFSPATYACRLASHCLAYANSCLNPLVYALASRH 297

Qy    404 ---------------HREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVG 447
Db    298 FRARFRRLWPCGRRRRHRARRA-LRRVRPASSG-----------PPGCPGDARPSGRLL 344

Qy    448 AGGSGLLPPP 457
Db    345 AGG-GQGPEP 353
```

RESULT 12
NQOC_THETH
ID NQOC_THETH STANDARD; PRT; 606 AA.
AC Q56227;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 12 (EC 1.6.5.3) (NADH
DE DEHYDROGENASE 1, CHAIN 12) (NDH-1, CHAIN 12).
GN NQO12.
OS Thermus aquaticus (subsp. thermophilus).
OC Bacteria; Thermus/Deinococcus group; Thermus group; Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB8 / ATCC 27634;
RX MEDLINE=97172490; PubMed=9020134;
RA Yano T., Chu S.S., Sled V.D., Ohnishi T., Yagi T.;
RT "The proton-translocating NADH-quinone oxidoreductase (NDH-1) of
RT thermophilic bacterium Thermus thermophilus HB-8. Complete DNA
RT sequence of the gene cluster and thermostable properties of the
RT expressed NQO2 subunit.";
RL J. Biol. Chem. 272:4201-4211(1997).
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC -!- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS. SUBUNITS NQO7-14
CC CONSTITUTE THE MEMBRANE SECTOR OF THE COMPLEX.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
CC OF CHLOROPLASTS OR MITOCHONDRIA.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; U52917; AAA97949.1; -.
DR InterPro; IPR003916; NADHub_oxdrdctse5.
DR InterPro; IPR001750; Oxidored_q1.
DR InterPro; IPR001516; Oxidored_q1_N.
DR Pfam; PF00361; oxidored_q1; 1.
DR Pfam; PF00662; oxidored_q1_N; 1.
KW Oxidoreductase; NAD; Ubiquinone; Transmembrane.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 29 49 POTENTIAL.
FT TRANSMEM 75 95 POTENTIAL.
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 126 146 POTENTIAL.
FT TRANSMEM 165 185 POTENTIAL.
FT TRANSMEM 206 226 POTENTIAL.
FT TRANSMEM 234 254 POTENTIAL.
FT TRANSMEM 266 286 POTENTIAL.
FT TRANSMEM 303 323 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT TRANSMEM 364 384 POTENTIAL.
FT TRANSMEM 397 417 POTENTIAL.
FT TRANSMEM 441 461 POTENTIAL.
FT TRANSMEM 492 512 POTENTIAL.
FT TRANSMEM 584 604 POTENTIAL.
SQ SEQUENCE 606 AA; 65141 MW; 47318E8F2665295C CRC64;

Query Match 3.8%; Score 110; DB 1; Length 606;
Best Local Similarity 21.4%; Pred. No. 1;
Matches 126; Conservative 69; Mismatches 198; Indels 196; Gaps 27

```
Qy     62 VLGLVCVPLLGSASDHWRGRYGR--RRPFIWALSLGILLSLFLIP----------RAGW 108
Db      3 LLGTILLPLLGFA---LLGLFGKRMREPLPGVLASGLVLASFLLGAGLLLSGGARFQAEW 59

Qy    109 LAGL---LCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVY- 164
Db     60 LPGIPFSLLLDNLSGFMLLIVTGVGFLIHVYAIGY---------MGGDPGYSR--FFAYF 108

Qy    165 ----AFMISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFG------------------- 201
Db    109 NLFIAMMLTLVLADSYPVMFIGWEGVGLASFL------LIGFWYKNPQYADSARKAFIVN 162

Qy    202 -------LLTLIFLTCVAATLLVAE--EAALGPTEPAEGLSAPSLSPHCCPCRARLAFRN 252
Db    163 RIGDLGFMLGMAILWALYGTLSISELKEAMEGP------LKNPDL----------LALAG 206

Qy    253 LGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEG----LYQGVPRAE 308
Db    207 L-----------------------------------LLFLGAVGKSAQIPLMVWLPDAM 230

Qy    309 PGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAA-G 367
Db    231 AGPTPVSALIHAATMVTAGVYLIARSSFLYSVLPD---VSYAIAVVGLLTAAYGALSAFG 287

Qy    368 ATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVF----------------- 410
Db    288 QTDIKKIVAYSTISQ--LGYMFLAAGVGAYWVA-LFHVFTHAFFKALLFLASGSVIHALG 344

Qy    411 --------------LPKYRGDT------------GGASSEDSLMTSFLPGPKPGAPFPN 443
Db    345 GEQDVRKMGGLWKHLPQTRWHALIGALALGGLPLLSGFWSKDAILAATLTYPFGGVGF-- 402

Qy    444 GHVGAGGSGLLPPPPALCGASACDVSVRVVVGE------PTEARVV---PGRGICLDLAI 494
Db    403 -YVGA----LL--VAVLTAMYAMRWFVLVFLGEERGHHHPHEAPPVMLWPNHLLALG-SV 454

Qy    495 LDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVS-AAGLGLVAIYFATQVV 542
Db    455 LAGYLALPHPLPNVLEPFLKPALAEVEAHHLSLGAEWGLIALSAAVALL 503
```

RESULT 13
YPUM_RHOCA
ID YPUM_RHOCA STANDARD; PRT; 477 AA.
AC P26176;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)

```
DE   HYPOTHETICAL 50.4 KDA PROTEIN IN PUHA-BCHM INTERGENIC REGION (ORF477)
DE   (PROTEIN F1696).
OS   Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1061;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84259352; PubMed=6744416;
RA   Youvan D.C., Bylina E.J., Alberti M., Begusch H., Hearst J.E.;
RT   "Nucleotide and deduced polypeptide sequences of the photosynthetic
RT   reaction-center, B870 antenna, and flanking polypeptides from R.
RT   capsulata.";
RL   Cell 37:949-957(1984).
CC   -!- SIMILARITY: TO RHODOBACTER CAPSULATUS PUCC AND TO
CC       BACTERIOCHLOROPHYLL SYNTHASE 44.5 KDA CHAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z11165; CAA77521.1; -.
DR   EMBL; K01183; -; NOT_ANNOTATED_CDS.
DR   PIR; B28988; B28988.
KW   Photosynthesis; Hypothetical protein.
SQ   SEQUENCE   477 AA;  50372 MW;  09A6A50AECD4D998 CRC64;


  Query Match             3.8%;  Score 109;  DB 1;  Length 477;
  Best Local Similarity   21.9%;  Pred. No. 0.96;
  Matches  122;  Conservative   80;  Mismatches  178;  Indels  176;  Gaps   29;

Qy      17 QLLLVNL--LTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSA 74
           ||| ::|  :: |:   |  | |.  ::||:||    :  :: | |||      :||
Db      30 QLLRLSLFQVSVGMAQVLLLG-TLNRVMILELGVPALVVAAMISI-PVLVAPFRAILGHR 87

Qy      75 SDHWRGRYG-RRRPFIWALSLGILLSLFLIPRA-----------GW----LAGLLCPDP 117
           || :|   | :| |::|  ||  :  | |:| :           |     ||
Db      88 SDTYRSALGWKRVPYLWFGSLWQMGGLALMPFSLILLSGDQTMGPAWAGEAFAG------ 141

Qy     118 RPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYA-FMISLG---GC 173
               :| |: |||:      :  |   || :|   :    :    :|  |:| :|
Db     142 ----VAFLMAGVGM-----HMTQTAGLALAADRATEETRPQVVALLYVMFLIGMGISAVI 192

Qy     174 LGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLS 233
           :|:||   |:|   |   :   : |  | :||: |  :|    : ::  : |   ||
Db     193 VGWLLR--DFDQITL---IRVVQGC--GAMTLV-LNVIA----LWKQEVMRPMTKAE-RE 239

Qy     234 APSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYT 293
           ||          | :||    ||                  ||  :   | |
Db     240 AP----------RQSFREAWGLL-----------------AAETGALRLLAT------ 264

Qy     294 DFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRA 353
                                  | :|:|   :|           | |:: :| :
Db     265 ---------------------------VMVGTLAFSMQ----------DVLLEPYGGQV 286

Qy     354 VYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPK 413
           : |       : ||            |  || || :|| ::   :  :: ||
Db     287 LGLKVGQTTWLTAG----------WAFGALVGFIWSARRL---SQGAVAHRVAA----- 327

Qy     414 YRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAG-GSGLLPPPPALCGASACDVSVRV 472
            ||   |  :  ::: | | | |    |    :| | |||:        | :   |:: |
Db     328 -RGLLVGIVAFTAVLFSPLFGSK--VLFFASAMGIGLGSGMF-------GIATLTVAMMV 377

Qy     473 VV------------GEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQS 519
           ||              : | |    |  : :  |  |   |::  | : ::||:   :
Db     378 VVRGASGIALGAWGAAQATAA----GLAVFIGGATRD---LVAHAAAAGYLGSLHSPALG 430

Qy     520 VTAYMVSAAGLGLVAI 535
```

```
           |   |:  ||  : :
Db     431 YTVVYVTEIGLLFITL 446


RESULT  14
CMLA_PSEAE
ID   CMLA_PSEAE     STANDARD;      PRT;   419 AA.
AC   P32482; Q56411; Q56412;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   CHLORAMPHENICOL RESISTANCE PROTEIN.
GN   CMLA.
OS   Pseudomonas aeruginosa.
OG   Plasmid R1033.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=Tn1696;
RX   MEDLINE=91294195; PubMed=1648560;
RA   Bissonnette L., Champetier S., Buisson J.-P., Roy P.H.;
RT   "Characterization of the nonenzymatic chloramphenicol resistance
RT   (cmlA) gene of the In4 integron of Tn1696: similarity of the product
RT   to transmembrane transport proteins.";
RL   J. Bacteriol. 173:4493-4502(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=Tn1696;
RX   MEDLINE=92052679; PubMed=1658833;
RA   Stokes H.W., Hall R.M.;
RT   "Sequence analysis of the inducible chloramphenicol resistance
RT   determinant in the Tn1696 integron suggests regulation by
RT   translational attenuation.";
RL   Plasmid 26:10-19(1991).
RN   [3]
RP   REVISIONS.
RC   TRANSPOSON=Tn1696;
RA   Partridge S., Hall R.M.;
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: IT APPEARS TO PROVOKE A REDUCTION OF THE CONTENT OF
CC       THE MAJOR PORINS OMPA AND OMPC.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC       AS THE DRUG RESISTANCE TRANSLOCASE FAMILY). BELONGS TO THE
CC       BCR/CMLA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U12338; AAB60004.1; -.
DR   EMBL; M64556; AAA26057.1; -.
DR   EMBL; M60454; AAA27462.1; -.
DR   EMBL; M60454; AAA27463.1; -.
DR   PIR; A47033; A47033.
KW   Antibiotic resistance; Transport; Transmembrane; Inner membrane;
KW   Plasmid; Transposable element.
FT   DOMAIN        1      9       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     10     30       1 (POTENTIAL).
FT   DOMAIN       31     46       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     47     67       2 (POTENTIAL).
FT   DOMAIN       68     80       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     81    101       3 (POTENTIAL).
FT   DOMAIN      102    109       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    110    127       4 (POTENTIAL).
FT   DOMAIN      128    141       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    142    162       5 (POTENTIAL).
```

```
FT   DOMAIN       163    169       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     170    187       6 (POTENTIAL).
FT   DOMAIN       188    214       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     215    235       7 (POTENTIAL).
FT   DOMAIN       236    245       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     246    266       8 (POTENTIAL).
FT   DOMAIN       267    279       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     280    299       9 (POTENTIAL).
FT   DOMAIN       300    306       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     307    327       10 (POTENTIAL).
FT   DOMAIN       328    342       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     343    363       11 (POTENTIAL).
FT   DOMAIN       364    370       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     371    391       12 (POTENTIAL).
FT   DOMAIN       392    419       EXTRACELLULAR (POTENTIAL).
SQ   SEQUENCE   419 AA;  44243 MW;  B5BF1D97DD7B9798 CRC64;


  Query Match             3.7%;  Score 107;  DB 1;  Length 419;
  Best Local Similarity   20.2%;  Pred. No. 1.2;
  Matches  102;  Conservative   73;  Mismatches  179;  Indels  150;  Gaps   23;

Qy      16 AQLLLVN----LLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVL-------GIGPVLG 64
           | :||::    | : |:::  | | : ::|  |   |      : : |       | | :
Db      14 ATVLLLSPFDLLASLGMDMYLPA-VPFMPNAL---GTTASTIQLTLTTYLVMIGAGQL-- 67

Qy      65 LVCVPLLGSASDHWRGRYGRRRPFIW------ALSLGILLSLFLIPRAGWLAGLLCPDPR 118
             | |  ||    | |||    :          |:|: |:
Db      68 -----LFGPLSD----RLGRRPVLLGGGLAYVVASMGLALT------------------- 99

Qy     119 PLELALLILGVGLLDFCG-QVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYL 177
             | : ||: :|  ||   |     | : |::   :     | :   |:::   :| |
Db     100 --SSAEVFLGLRILQACGASACLVSTFATVRDIYAGREESNVIYGILGSMLAMVPAVGPL 157

Qy     178 LPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
           | |:      :  :|| :   :|  | |  :   ||      |  :
Db     158 LGAL------VDMWLGWR--AIFAFLGLGMIAASAAAWRFWPETRV-------------- 195

Qy     238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
                 |:|      ||         :|      ::  ||    : :| :|::  :
Db     196 --------QRVAGLQWSQLL---------LPVKCLNFWLYTLCYAAGMGSFFVFFS--IA 236

Qy     298 EGLYQGVPRAEPGTEARRHYDEGV-RMGSLGLFLQCAISLVFSL-VMDRLVQRFGTRAVY 355
           ||  |               :|| ::|   ||   ||::||:   | |:: ::|: :|
Db     237 PGLMMG-------------RQGVSQLGFSLLFATVAIAMVFTARFMGRVIPKWGSPSVL 282

Qy     356 LASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYR 415
             :    : |||  |  ::  : |  :: ||      | |  |  :
Db     283 RMGMGC--LIAGAVLL--AITEIWALQSVLGF------IAPMWLVGI------------- 319

Qy     416 GDTGGASSEDSLMTS-------FLPGPKPGAPFPNGHVGAGGSG-----LLPPPPALCGA 463
             | ::  |:  :       : |      |  | |  |  |      |||   |
Db     320 ----GVATAVSVAPNGALRGFDHVAGTVTAVYFCLGGVLLGSIGTLIISLLPRNTAWPVV 375

Qy     464 SACDVSVRVVVGEPTEARVVPGRG 487
             |     ||:|    :||   ||
Db     376 VYCLTLATVVLGLSCVSRVKGSRG 399


RESULT  15
YJ94_YEAST
ID   YJ94_YEAST     STANDARD;      PRT;   448 AA.
AC   P47159;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   HYPOTHETICAL 49.7 KDA PROTEIN IN RPS5-ZMS1 INTERGENIC REGION.
GN   YJR124C OR J2046.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rose M., Koetter P., Entian K.D.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboratio
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49624; CAA89655.1; -.
DR   SGD; S0003885; YJR124C.
KW   Hypothetical protein; Transmembrane; ATP-binding.
FT   TRANSMEM      51     71       POTENTIAL.
FT   TRANSMEM      94    114       POTENTIAL.
FT   TRANSMEM     147    167       POTENTIAL.
FT   TRANSMEM     185    205       POTENTIAL.
FT   TRANSMEM     261    281       POTENTIAL.
FT   TRANSMEM     288    308       POTENTIAL.
FT   TRANSMEM     334    354       POTENTIAL.
FT   TRANSMEM     387    407       POTENTIAL.
FT   TRANSMEM     417    437       POTENTIAL.
FT   NP_BIND      386    393       ATP (POTENTIAL).
SQ   SEQUENCE   448 AA;  49663 MW;  2319CF9DD1220C14 CRC64;


  Query Match             3.7%;  Score 105;  DB 1;  Length 448;
  Best Local Similarity   21.8%;  Pred. No. 1.7;
  Matches   94;  Conservative   62;  Mismatches  163;  Indels  112;  Gaps   25;

Qy       2 VQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEV--GVEEK---FMTMV 56
           :: || |  ||          ||::||          |  | |      |:|   ||::
Db      16 IKLLWASVFLR----------LLSYGL-------TNQVLTLFLNAINMTEDKIGLFMSLT 58

Qy      57 LGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPD 116
           |  | |   :|  :|   :| |    ||||  ::  :: :            |:||:
Db      59 LA-GDV---ICSYILTWYADSW----GRRRVLVYGCAMML-----------LSGLVFSF 98

Qy     117 PRPLELALLILGVGLL----DFCGQVCFTPL-EALLSDLFRDPDHCR-QAYSVYAFMISL 170
             | |:    |::    |  |   |  : ||::: |   |  : | : |:::| : ::
Db      99 SENFTLLLVFAIFGVISPSSDEVGP--FKSIEEAMIAHL--SPHNARPEIYAIHALVGTI 154

Qy     171 GGCLGYLLPAIDWDTSALAPYLGTQEEC--LFGLLTLIFLTCVAATLLVAEEAA-----L 223
           |  || ::  |  |          |  :|  |  ||   |  |    :|:  :|
Db     155 GSALGAIICGIFVDLLKRTGLAATDLQCYKLVFLLYAFFAFCKMVIMLLLSDATELDGHY 214

Qy     224 GPTEPAEGLSAP-SLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSW 282
             |:  |  : |  ::    |   :          | ::|       | :| |
Db     215 EHTDCNEETAEPLDVNDETAPLMRQATHPE-----ERSNKLSKETVSVLMKLLV------ 263

Qy     283 MALMTFTLFYTDFVGEG-------LYQGVPRAEPGTEARRHYDEGVRMGSLG---LFLQC 332
                 :|  | :| |        :|               :| :   ||||    ||
Db     264 -------IFMVDSLGSGFMTSGWMVY--------------YYSKQFLMGSLALGTLFFIT 302

Qy     333 AISLVFSLVMDRLVQR-FG-TRAVYLASVAA------FPVAAGATCLSHSVAVVTASAAL 384
             : :  | :   ::  | ||  ||  |  : :        |:|     |  |:  :    |
Db     303 QLVMASSTIPSSIIARCFGPVRATLLVQIPSGIFSILIPMA--KNYLPLSILFLNLHFAT 360

Qy     385 TGFTFSALQIL 395
           |    :  |||
Db     361 TAMDVTPRQIL 371


Search completed: February 26, 2002, 07:58:17
Job time: 102 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 07:54:50 ; Search time 19.5 Seconds
(without alignments)
2160.232 Million cell updates/sec

Title: US-09-605-783A-113
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	347.5	12.1	515	2	T14340	sucrose-proton tra
2	337.5	11.8	523	2	T12198	sucrose transport
3	330	11.5	516	2	JQ2389	sucrose transport
4	328.5	11.5	594	2	G84441	probable sucrose-p
5	324.5	11.3	525	2	S28052	sucrose transport
6	323.5	11.3	512	2	S38196	sucrose transport
7	323.5	11.3	512	2	G86360	hypothetical prote
8	317.5	11.1	507	2	S48789	sucrose transport
9	311	10.9	533	2	S43142	sucrose transport
10	305.5	10.7	501	2	T14339	sucrose-proton tra
11	297.5	10.4	428	2	S48788	sucrose transport
12	294	10.3	510	2	S38657	sucrose transport
13	294	10.3	523	2	S51114	sucrose-proton sym
14	292.5	10.2	512	2	F96741	probable sucrose t
15	292	10.2	537	2	T02982	probable sucrose t
16	286	10.0	474	2	A86234	hypothetical prote
17	281.5	9.8	491	2	E96691	probable sucrose-p
18	279.5	9.8	513	2	S38197	sucrose transport
19	276.5	9.7	492	2	A84520	probable sucrose-p
20	273	9.5	503	2	S52377	sucrose transport
21	235.5	8.2	553	2	T38541	probable sucrose c
22	191	6.7	452	2	F75217	hypothetical prote
23	142	5.0	544	2	S75696	melibiose carrier
24	139.5	4.9	454	2	A75444	hypothetical prote
25	138.5	4.8	430	2	E75217	transporter PAB217
26	133	4.6	389	2	G83413	probable MFS trans
27	128.5	4.5	422	2	G83503	probable MFS trans
28	128.5	4.5	594	2	A83096	probable permease
29	127.5	4.5	472	2	F82639	resistance protein
30	127	4.4	451	2	E81781	probable integral
31	126	4.4	451	2	A81206	sugar transporter,
32	125.5	4.4	399	2	JQ1479	tetracycline resis
33	125	4.4	391	2	S74688	hypothetical prote
34	125	4.4	407	2	H75515	tetracycline-efflu
35	124.5	4.4	399	1	S38656	tetA protein - Pse
36	121.5	4.2	503	2	S43017	puromycin resistan
37	120.5	4.2	414	2	F83119	probable transport
38	120.5	4.2	466	2	A86033	probable permease
39	119.5	4.2	399	1	YTECR1	tetracycline resis
40	119.5	4.2	491	2	C85540	regulates beta-lac
41	119.5	4.2	491	2	S37391	signal transducer
42	119	4.2	409	2	F83493	probable MFS trans
43	119	4.2	440	2	D86025	probable transport
44	117.5	4.1	403	2	E84063	multidrug resistan
45	117.5	4.1	448	2	H83335	probable MFS trans

ALIGNMENTS

RESULT 1
T14340
sucrose-proton transport protein - carrot
N;Alternate names: sucrose/H+ symporter protein
C;Species: Daucus carota (carrot)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T14340
R;Shakya, R.; Sturm, A.
Plant Physiol. 118, 1473-1480, 1998
A;Title: Characterization of source- and sink-specific sucrose/H+ symporters from
A;Reference number: Z17991; MUID:99063785
A;Accession: T14340
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-515 <SHA>
A;Cross-references: EMBL:Y16768; NID:g2969883; PIDN:CAA76369.1; PID:g2969884
A;Experimental source: cultivar Namtaise; root
C;Genetics:
A;Note: SUT2
C;Superfamily: common tobacco sucrose transport protein

Query Match 12.1%; Score 347.5; DB 2; Length 515;
Best Local Similarity 25.9%; Pred. No. 9.3e-19;
Matches 131; Conservative 80; Mismatches 215; Indels 79; Gaps 17;

```
Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|:||  :  |::  |  :: : | :  :|:  |:   :   ||: |::  |::|  ||
Db      34 KLVLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 93

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
           | :  :|||||| : :  : :|: ||   |: | :       :    :|  : : ::|
Db      94 HCQSSFGRRRPFIASGAGCVAISVILI---GFAADISYKAGDDMSKTLKPRAVTVFVIGF 150

Qy     130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LL 178
           :||    :  |  |||:||   |    | | : |:| :::|  |||          |
Db     151 WILDVANNMLQGPCRALLADLCSGDTRRMRSANAFYSFFMAVGNILGYAAGSYNNLYKLF 210

Qy     179 PAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLS 238
           |    | |  |    : |    : |: :  | |  :| | :  || : |:    |
Db     211 PFS--KTHACDLYCANLKSCFIISIALLIIITVVALSVVRENS--GPPDDADAAEEP--- 263

Qy     239 PHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGE 298
           |   ::    | ||  |  |    || :  | :    :|:|   | || ||::|
Db     264 ----PSSGKIPV--FGELLGALKDL----PRPMLLLLIVTCLNWIAWFPFILFDTDWMGR 313

Qy     299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
           :| |   |  |    : ||:||| |:||| |   :  : |: :: ||:  |   : |
Db     314 EIYGGT--AGQG----KLYDQGVRAGALGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LWG 366

Qy     359 VAAFPVAAGATCL----------SHSV--------AVVTASA----ALTGFTFSALQIL 395
             | :| |               ||         | | | |    :: |   |    :
```

```
Db    367 FVNFILAIGLVMTVVVSKVAQHQREHSANGQLLPPSAGVKAGALSLFSILGIPLSITYSI 426

Qy    396 PYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPG-------APFPNGHVGA 448
Db    427 PFALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAFVVGAISA 486

Qy    449 GGSGLLP----PPPALCGASACDVS 469
Db    487 AISGVLAIVLLPKPSKDAASKLSLS 511


RESULT   2
T12198
sucrose transport protein - fava bean
C;Species: Vicia faba (fava bean)
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 20-Jun-2000
C;Accession: T12198
R;Weber, H.; Borisjuk, L.; Heim, U.; Sauer, N.; Wobus, U.
Plant Cell 9, 895-908, 1997
A;Title: A role for sugar transporters during seed development: molecular characterizati
A;Reference number: Z17451; MUID:97355984
A;Accession: T12198
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-523 <WEB>
A;Cross-references: EMBL:Z93774
C;Genetics:
A;Gene: sut
C;Superfamily: common tobacco sucrose transport protein


  Query Match             11.8%;  Score 337.5;  DB 2;  Length 523;
  Best Local Similarity   25.8%;  Pred. No. 5.4e-18;
  Matches  134;  Conservative   76;  Mismatches  185;  Indels  125;  Gaps   20;

Qy     17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
Db     36 KIMVVASIAAGVQFGWALQLSLLTPYVQLLGIHHTWAAYIWLCGPISGMLVQPIVGYHSD 95

Qy     77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
Db     96 RCTSRFGRRRPFIAAGSIAVAIAVFLI---GYAADLGHSFGDSLDQKVRPRAIGIFVVGF 152

Qy    130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID------ 182
Db    153 WILDVANNMLQGPCRALLGDLCAGNQRKTRNANAFFSFFMAVGNVLGYAAGAYSKLYHVF 212

Qy    183 --WDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLV-AEEAALGPTEPA----EGLSAP 235
Db    213 PFTKTKACNVYCANLKSCFF--LSIALLTVLATSALIYVKETALTPEKTVVTTEDGGSSG 270

Qy    236 SLSPHCCPCRARL--AFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYT 293
Db    271 GM-----PCFGQLSGAFKEL-------------KRPMWILLLVTCLNWIAWFPFLLFDT 311

Qy    294 DFVGEGLYQGVPRAEPGTEARRH-YDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTR 352
Db    312 DWMGKEVY-------GGTVGEGHAYDMGVREGALGLMLNSVVLGATSLGVDILARGVG-G 363

Qy    353 AVYLASVAAFPVAAGATCLSHSVAVV-------------------------TASAALTGF 387
Db    364 VKRLWGIVNFLL---AICLGLTVLVTKLAQHSRQYAPGTGALGDPLPPSEGIKAGALTLF 420

Qy    388 TFSALQI-----LPYTLASLYHREKQVFLPKYRGDTGGASSEDSL------------MTS 430
Db    421 SVLGVPLAITYSIPFALASIF------------SSTSGAGQGLSLGVLNLAIVIPQMFVS 468

Qy    431 FLPGPKPG-------APFPNGHVGAGGSG-----LLPPPP 458
Db    469 VLSGPWDALFGGGNLPAFVVGAVAALASGILSIILLPSPP 508


RESULT   3
```

```
JQ2389
sucrose transport protein - potato
C;Species: Solanum tuberosum (potato)
C;Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 21-Jul-2000
C;Accession: JQ2389; S40310
R;Riesmeier, J.W.; Hirner, B.; Frommer, W.B.
Plant Cell 5, 1591-1598, 1993
A;Title: Potato sucrose transporter expression in minor veins indicates a role in p
A;Reference number: JQ2389; MUID:94146554
A;Accession: JQ2389
A;Molecule type: mRNA
A;Residues: 1-516 <RIE>
A;Cross-references: EMBL:X69165; NID:g439293; PIDN:CAA48915.1; PID:g439294
A;Experimental source: cv. Desiree
C;Comment: The gene encoding for this protein is highly expressed in mature leaves.
C;Superfamily: common tobacco sucrose transport protein
C;Keywords: glycoprotein; transmembrane protein
F;31-53/Domain: transmembrane #status predicted <TM1>
F;67-86/Domain: transmembrane #status predicted <TM2>
F;103-122/Domain: transmembrane #status predicted <TM3>
F;141-160/Domain: transmembrane #status predicted <TM4>
F;180-200/Domain: transmembrane #status predicted <TM5>
F;226-248/Domain: transmembrane #status predicted <TM6>
F;285-304/Domain: transmembrane #status predicted <TM7>
F;331-349/Domain: transmembrane #status predicted <TM8>
F;366-385/Domain: transmembrane #status predicted <TM9>
F;409-427/Domain: transmembrane #status predicted <TM10>
F;429-448/Domain: transmembrane #status predicted <TM11>
F;3,92/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             11.5%;  Score 330;  DB 2;  Length 516;
  Best Local Similarity   24.7%;  Pred. No. 2e-17;
  Matches  132;  Conservative   93;  Mismatches  197;  Indels  112;  Gaps   21;

Qy      4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
Db     29 KLW----------KIIVVASIAAGVQFGWALQLSLLTPYVQLLGIPHKFASFIWLCGPIS 78

Qy     64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPD 116
Db     79 GMIVQPVVGYYSDNCSSRFGRRRPFIAAGAALVMIAVFLI---GFAADLGHASGDTLGKG 135

Qy    117 PRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPD-HCRQAYSVYAFMISLGGCLG 175
Db    136 FKPRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKSGRMRTANAFFSFFMAVGNILG 195

Qy    176 YLLPAIDW--------DTSALAPYLGTQEECLF-GLLTLIFLTCVAATLLVAEE-AALGP 225
Db    196 YAAGSYSHLFKVFPFSKTKACDMYCANLKSCFFIAIFLLLSLTTIALTLVRENELPEKDE 255

Qy    226 TEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMAL 285
Db    256 QEIDEKLAGAG--------KSKVPF--FGEIFGALKEL----PRPMWILLLVTCLNWIAW 301

Qy    286 MTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRL 345
Db    302 FPFFLYDTDWMAKEVFGG----QVGD--ARLYDLGVRAGAMGLLLQSVVLGFMSLGVEFL 355

Qy    346 VQRFGTRAVYLASVAAFPVAAGATCLSHSVAVV---------TASAALTGFT-------- 388
Db    356 GKKIG-GAKRLWGILNFVL---AICLAMTILVTKMAEKSRQHDPAGTLMGPTPGVKIGAL 411

Qy    389 --FSALQI-------LPYTLASLYHREK------------QVFLPKYRGDTGGASSEDS 426
Db    412 LLFAALGIPLAATFSIPFALASIFSSNRGSGQGLSLGVLNLAIVVPQMLVSLVGGPWDDL 471

Qy    427 LMTSFLPGPKPGAPFPNGHVGAGGSG-----LLPPPPALCGASACDVSVRVVVG 475
Db    472 FGGGNLPG------FVVGAVAAAASAVLALTMLPSPPA-------DAKPAVAMG 512


RESULT   4
```

```
G84441
probable sucrose-proton symporter [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: G84441
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: G84441
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-594 <STO>
A;Cross-references: GB:AE002093; NID:g3461813; PIDN:AAC32907.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g02860
A;Map position: 2
C;Superfamily: common tobacco sucrose transport protein


  Query Match             11.5%;  Score 328.5;  DB 2;  Length 594;
  Best Local Similarity   24.1%;  Pred. No. 2.9e-17;
  Matches  120;  Conservative   72;  Mismatches  174;  Indels  131;  Gaps   18;

Qy      22 NLLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGS 73
           :|:|  |   :|||: :         : | :  :|:   | : :   ||: |||  | :|
Db      59 SLVTLVLSCTVAAGVQFGWALQLSLLTPYIQTLGISHAFSSFIWLCGPITGLVVQPFVGI 118

Qy      74 ASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPD----------PRPLE 121
            ||    :||||||||   |  | ::: :|   |: |  | |  |           |
Db     119 WSDKCTSKYGRRRPFILVGSFMISIAVIII---GFSADIGYLLGDSKEHCSTFKGTRTRA 175

Qy     122 LALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAI 181
             : |:|  |||         |  |||:||   ||    | :|:   :::|  ||:   |
Db     176 AVVFIIGFWLLDLANNTVQGPARALLADL-SGPDQRNTANAVFCLWMAIGNILGFSAGAS 234

Qy     182 ----DWDTSALAPYLGTQEECLFG--------LLTLIFLT-CVAATLLVAEEAALGPTEP 228
               :|      |:| |   |            || ::|||  |   |:  |:|      :|
Db     235 GKWQEW-----FPFL-TSRACCAACGNLKAAFLLAVVFLTICTLVTIYFAKEIPFTSNKP 288

Qy     229 AE-GLSAPSLSPHCCPCRARLAFRNL---------------------------------- 253
               ||| |          |  : |
Db     289 TRIQDSAPLLDD--------LQSKGLEHSKLNNGTANGIKYERVERDTDEQFGNSENEHQ 340

Qy     254 --------GALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVP 305
                   |::|  |    :|  :  : :    :|::  | || ||::|  :| | |
Db     341 DETYVDGPGSVLVNLLTSLRHLPPAMHSVLIVMALTWLSWFPFFLFDTDWMGREVYHGDP 400

Qy     306 RAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVY-LASVAAFPV 364
           :  :     ||:||| |:||| |   :  : | ::: : || | | |: |::   |
Db     401 TGD--SLHMELYDQGVREGALGLLLNSVVLGISSFLIEPMCQRMGARVVWALSNFTVFAC 458

Qy     365 AAGATCLS--------------------HSVAVVTASAALTGFTFSALQILPYTLASLY 403
             ||   :|                    : ||:    || ||  :    :|::: :
Db     459 MAGTAVISLMSLSDDKNGIEYIMRGNETTRTAAVIV--FALLGFPLAITYSVPFSVTA-- 514

Qy     404 HREKQVFLPKYRGDTGG 420
                 :   |:||
Db     515 ---------EVTADSGG 522


RESULT   5
S28052
sucrose transport protein - spinach
N;Alternate names: sucrose carrier protein; sucrose permease
C;Species: Spinacia oleracea (spinach)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 21-Jul-2000
C;Accession: S28052
R;Riesmeier, J.W.; Willmitzer, L.; Frommer, W.B.
EMBO J. 11, 4705-4713, 1992
A;Title: Isolation and characterization of a sucrose carrier cDNA from spinach by functi
A;Reference number: S28052; MUID:93099843
A;Accession: S28052
A;Molecule type: mRNA
A;Residues: 1-525 <RIE>
A;Cross-references: EMBL:X67125; NID:g21318; PIDN:CAA47604.1; PID:g21319
C;Superfamily: common tobacco sucrose transport protein
C;Keywords: transmembrane protein


  Query Match             11.3%;  Score 324.5;  DB 2;  Length 525;
  Best Local Similarity   23.2%;  Pred. No. 5.2e-17;
  Matches  129;  Conservative   91;  Mismatches  196;  Indels  141;  Gaps   20;

Qy      23 LLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSA 74
           |   ||   :|||: :         : | :  :|:   :   :   ||: |::  ||:|
Db      35 LKKLGLVASVAAGVQFGWALQLSLLTPYVQLLGIPHTWAAYIWLCGPISGMIVQPLVGYY 94

Qy      75 SDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPDP-----RPLELALLIL 127
           ||    |:|||||| | :  : ::: ||   |: | |    ||     :|  :|: ::
Db      95 SDRCTSRFGRRRPFIAAGAALVAVAVGLI---GFAADIGAASGDPTGNVAKPRAIAVFVV 151

Qy     128 GVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID---- 182
           |  :||        |  |||:|:          | | : ::| ::||   ||   :
Db     152 GFWILDVANNTLQGPCRALLADMAAGSQTKTRYANAFFSFFMALGNIGGYAAGSYSRLYT 211

Qy     183 ----WDTSALAPYLGTQEECLFGLLT-LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
                 |:|   |    : | |  :| || || :| :::   :  :   :  | |   :
Db     212 VFPFTKTAACDVYCANLKSCFFISITLLIVLTILALSVVKERQITIDEIQEEEDLKNRNN 271

Qy     238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
           |  |    ||| |   | |:  |  |    |: :  | :    :|:|   | || ||::|
Db     272 SSGC----ARLPF--FGQLIGALKDL----PKPMLILLLVTALNWIAWFPFLLFDTDWMG 321

Qy     298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG--TRAVY 355
           : :| |           : ||:||  |:||| :   :  | || :: | :  |   |
Db     322 KEVYGGT------VGEGKLYDQGVHAGALGLMINSVVLGVMSLSIEGLARMVGGAKRLWG 375

Qy     356 LASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYR 415
           : ::        | ||: :| :|| ||                 : |:    :
Db     376 IVNIIL------AVCLAMTV-LVTKSAE-----------------HFRDSHHIM---- 405

Qy     416 GDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVG 475
                                                || : |||||  |     :::  |:|
Db     406 --------------------------------GSAVPPPPPA--GVKGGALAIFAVLG 429

Qy     476 EPTEARV-VP-----------GRGICLDLAILDSAFLLSQV----------------AP 506
           |      :|           | |  | | :|: | :: |:                  |
Db     430 IPLAITFSIPFALASIFSASSGSGQGLSLGVLNLAIVVPQMFVSVTSGPWDAMFGGGNLP 489

Qy     507 SLFMGSIVQLSQSVTAY 523
           :  :|::   : :| ::
Db     490 AFVVGAVAATASAVLSF 506


RESULT   6
S38196
sucrose transport protein SUC2 - Arabidopsis thaliana
N;Alternate names: sucrose-proton symporter SUC2
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
C;Accession: S38196; T00773
R;Sauer, N.K.
submitted to the EMBL Data Library, October 1993
A;Description: SUC1 and SUC2: two sucrose transporters from Arabidopsis thalian
A;Reference number: S38196
A;Accession: S38196
A;Molecule type: mRNA
A;Residues: 1-512 <SAU>
A;Cross-references: EMBL:X75382; NID:g407091; PID:g407092
R;Vysotskaia, V.S.; Schwartz, J.R.; Toriumi, M.; Yu, G.; Oji, O.; Kwan, A.; Liu S
rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federsp
submitted to the EMBL Data Library, July 1998
A;Description: Arabidopsis thaliana chromosome 1 BAC T22J18 sequence, complete seq
```

```
A;Reference number: Z14202
A;Accession: T00773
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-490,'A',492-512 <VYS>
A;Cross-references: EMBL:AC003979; NID:g3172156; PID:g3287687; GSPDB:GN00059; ATSP:T22J1
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: SUC2; ATSP:T22J18.12
A;Map position: 1
A;Introns: 419/3; 441/1; 455/3
C;Superfamily: common tobacco sucrose transport protein
C;Keywords: sugar transport


  Query Match             11.3%;  Score 323.5;  DB 2;  Length 512;
  Best Local Similarity   24.1%;  Pred. No. 6e-17;
  Matches  123;  Conservative   92;  Mismatches  195;  Indels  101;  Gaps   18;

Qy      14 RKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGS 73
           |  ::: |: :  |::   |  :: : | :  :|:  |: :::   ||: |::  |::|
Db      28 RLRKIISVSSIAAGVQFGWALQLSLLTPYVQLLGIPHKWASLIWLCGPISGMLVQPIVGY 87

Qy      74 ASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLI 126
             ||    |:|||||||| | :  : :::|||   |: | :       |   |:   :|:
Db      88 HSDRCTSRFGRRRPFIVAGAGLVTVAVFLI---GYAADIGHSMGDQLDKPPKTRAIAIFA 144

Qy     127 LGVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY-------LL 178
           ||  :||         |  | |:||   :    | | : ::| :::|  |||       |
Db     145 LGFWILDVANNTLQGPCRAFLADLSAGNAKKTRTANAFFSFFMAVGNVLGYAAGSYRNLY 204

Qy     179 PAIDWD-TSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
             : :  | :   |    : | |  :||: |     :|   :|    |   |:| :
Db     205 KVVPFTMTESCDLYCANLKTCFFLSITLL-LIVTFVSLCYVKEKPWTPEPTADGKA---- 259

Qy     238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
                 : :  |   | :     :|     | :  | :    :|:|   | || ||::|
Db     260 --------SNVPF--FGEIFGAFKEL----KRPMWMLLIVTALNWIAWFPFLLFDTDWMG 305

Qy     298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLA 357
             :| |   |     ::: |::||| |:||| |   :    || :: : :: |  |  |
Db     306 REVYGGNSDATATAASKKLYNDGVRAGALGLMLNAIVLGFMSLGVEWIGRKLG-GAKRLW 364

Qy     358 SVAAFPVAAGATCLSHSVAV---------------------VTASA----ALTG----FT 388
             :  | :   | ||: :| |                     ||| |    |: |     |
Db     365 GIVNFIL---AICLAMTVVVTKQAENHRRDHGGAKTGPPGNVTAGALTLFAILGIPQAIT 421

Qy     389 FSALQILPYTLASLYHREK-------------QVFLPKYRGDTGGASSEDSLMTSFLPGP 435
           ||    :|: |||::                 : :|:     ||   ::      :|
Db     422 FS----IPFALASIFSTNSGAGQGLSLGVLNLAIVVPQMVISVGGGPFDELFGGGNIPA- 476

Qy     436 KPGAPFPNGHVGAGGSGLL-------PPPPA 459
                 |  | : |  ||:|       ||| |
Db     477 -----FVLGAIAAAVSGVLGLTVLPSPPPDA 502


RESULT   7
G86360
hypothetical protein AAC25515.1 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: G86360
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: G86360
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-512 <STO>
A;Cross-references: GB:AE005172; NID:g3287687; PIDN:AAC25515.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
C;Superfamily: common tobacco sucrose transport protein


  Query Match             11.3%;  Score 323.5;  DB 2;  Length 512;
  Best Local Similarity   24.1%;  Pred. No. 6e-17;
  Matches  123;  Conservative   92;  Mismatches  195;  Indels  101;  Gaps   18;

Qy      14 RKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGS 73
           |  ::: |: :  |::   |  :: : | :  :|:  |: :::   ||: |::  |::|
Db      28 RLRKIISVSSIAAGVQFGWALQLSLLTPYVQLLGIPHKWASLIWLCGPISGMLVQPIVGY 87

Qy      74 ASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLI 126
             ||    |:|||||||| | :  : :::|||   |: | :       |   |:   :|:
Db      88 HSDRCTSRFGRRRPFIVAGAGLVTVAVFLI---GYAADIGHSMGDQLDKPPKTRAIAIFA 144

Qy     127 LGVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY-------LL 178
           ||  :||         |  | |:||   :    | | : ::| :::|  |||       |
Db     145 LGFWILDVANNTLQGPCRAFLADLSAGNAKKTRTANAFFSFFMAVGNVLGYAAGSYRNLY 204

Qy     179 PAIDWD-TSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
             : :  | :   |    : | |  :||: |     :|   :|    |   |:| :
Db     205 KVVPFTMTESCDLYCANLKTCFFLSITLL-LIVTFVSLCYVKEKPWTPEPTADGKA---- 259

Qy     238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
                 : :  |   | :     :|     | :  | :    :|:|   | || ||::|
Db     260 --------SNVPF--FGEIFGAFKEL----KRPMWMLLIVTALNWIAWFPFLLFDTDWMG 305

Qy     298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLA 357
             :| |   |     ::: |::||| |:||| |   :    || :: : :: |  |  |
Db     306 REVYGGNSDATATAASKKLYNDGVRAGALGLMLNAIVLGFMSLGVEWIGRKLG-GAKRLW 364

Qy     358 SVAAFPVAAGATCLSHSVAV---------------------VTASA----ALTG----FT 388
             :  | :   | ||: :| |                     ||| |    |: |     |
Db     365 GIVNFIL---AICLAMTVVVTKQAENHRRDHGGAKTGPPGNVTAGALTLFAILGIPQAIT 421

Qy     389 FSALQILPYTLASLYHREK-------------QVFLPKYRGDTGGASSEDSLMTSFLPGP 435
           ||    :|: |||::                 : :|:     ||   ::      :|
Db     422 FS----IPFALASIFSTNSGAGQGLSLGVLNLAIVVPQMVISVGGGPFDELFGGGNIPA- 476

Qy     436 KPGAPFPNGHVGAGGSGLL-------PPPPA 459
                 |  | : |  ||:|       ||| |
Db     477 -----FVLGAIAAAVSGVLALTVLPSPPPDA 502


RESULT   8
S48789
sucrose transport protein - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 26-May-2000
C;Accession: S48789
R;Buerkle, X.Y.Z.; Frommer, W.B.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48787
A;Accession: S48789
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-507 <BUE>
A;Cross-references: EMBL:X82276; NID:g575350; PID:g575351
C;Superfamily: common tobacco sucrose transport protein


  Query Match             11.1%;  Score 317.5;  DB 2;  Length 507;
  Best Local Similarity   24.3%;  Pred. No. 1.7e-16;
  Matches  125;  Conservative   87;  Mismatches  203;  Indels   99;  Gaps   18;

Qy       4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
```

```
           :||          ::::|  :  |::   |  :: : | :  :|:  || : :   ||:
Db      23 KLW----------KIIMVASIAAGVQFGWALQLSLLTPYVQLLGIPHKFASFIWLCGPIS 72

Qy      64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDP-----R 118
           |::  |::|  ||:   |:|||| || | :  : :::|||  |  | |    ||     :
Db      73 GMIVQPVVGYYSDNCSSRFGRRRGFIAAGAALVTIAVFLIGFAADL-GHATGDPLGKGSK 131

Qy     119 PLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLL 178
           |  :|: ::|  :||    :   | |||:||       | : : ::| :::|  |||
Db     132 PRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKARMRTSNAFFSFFMAVGNVLGYAA 191

Qy     179 PAID--------WDTSALAPYLGTQEECLF-GLLTLIFLTCVAATLLVAEEAALGPTEPA 229
            :          | |  |   : | |  :  |: || :| |  |  |  |  |   :
Db     192 GSYSRLCKIFPFSKTPACDIYCANLKSCFFIAVFLLLSLTILALT--VVRENELPEKDEH 249

Qy     230 EGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFT 289
           |      :       :::: |   | :   |  |    || :  | :    :|:|   |
Db     250 E------IDEKAGARKSKVPF--FGEIFGALKDL----PRPMWILLLVTSLNWIARFPFF 297

Qy     290 LFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRF 349
           |: ||:: : :| |  :   |    | || ||  |:||| |   :    || :: | ::
Db     298 LYDTDWMAKEVYGG--KVGDG----RLYDLGVHAGALGLLLNSVVLGFMSLSVEFLGKKI 351

Qy     350 GTRAVYLASVAAFPVAAGATCLSHSVAVV------------TASAALTGFTFSALQI-- 394
           |     |  :  | :   | |:: :| |              |  |  :|    || :
Db     352 G-GVKRLWGILNFVL---AVCMALTVLVTKMAEKSRQYDAHGTLMAPTSGVKIGALTLFA 407

Qy     395 ----------LPYTLASLYHREK------------QVFLPKYRGDTGGASSEDSLMTS 430
                    :|: |||::                : :|:      |   :|
Db     408 VLGIPLAVTFSVPFALASIFSSNAGSGQGLSLGVLNLAIVVPQMLVSIAGGPWDDLFGGG 467

Qy     431 FLPGPKPGAPFPNGHVGAGGSG-----LLPPPPA 459
            |||      |  | | |  ||     :|| |||
Db     468 NLPG------FIVGAVAAAASGILALTMLPSPPA 495


RESULT  9
S43142
sucrose transport protein - castor bean
N;Alternate names: sucrose carrier
C;Species: Ricinus communis (castor bean)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 26-May-2000
C;Accession: S43142
R;Weig, A.; Komor, E.
submitted to the EMBL Data Library, March 1994
A;Description: A sucrose carrier from Ricinus communis.
A;Reference number: S43142
A;Accession: S43142
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-533 <WEI>
A;Cross-references: EMBL:Z31561; NID:g468561; PID:g468562
C;Superfamily: common tobacco sucrose transport protein


  Query Match             10.9%;  Score 311;  DB 2;  Length 533;
  Best Local Similarity   23.3%;  Pred. No. 5.5e-16;
  Matches  134;  Conservative   95;  Mismatches  225;  Indels  120;  Gaps   19;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           ::::|  :  |::   |  :: : | :  :|:   :   :   ||: |::  |::|  ||
Db      37 KVVMVASIAAGIQFGWALQLSLLTPYVQLLGIPHTWAAFIWLCGPISGMLVQPIVGYHSD 96

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRA---GWLAG-LLCPDPRPLELALLILGVGLL 132
              |:|||||| : :  : :::|||  |   | |:|  |   |:  :|: ::|  :|
Db      97 RCTSRFGRRRPFIASGAAFVAIAVFLIGYAADLGHLSGDSLDKSPKTRAIAIFVVGFWIL 156

Qy     133 DFCGQVCFTPLEALLSDLF-RDPDHCRQAYSVYAFMISLGGCLGYLLPAID--------W 183
           |    :  | |||:||       | | ::::| :::|  |||   |
Db     157 DVANNMLQGPCRALLADLSGTSQKKTRTANALFSFFMAVGNVLGYAAGAYTHLYKLFPFT 216

Qy     184 DTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGL--------SAP 235
           |:|  |   : | |  : |:    | |   | |:    |  | : :       |
```

```
Db     217 KTTACDVYCANLKSCFFISIVLLLSLTVLALSYVKEK----PWSPDQAVDNAEDDTASQA 272

Qy     236 SLSPHCCPCRARL--AFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYT 293
           | |    |    :  ||:||               | :  | :    :|:|   | || |
Db     273 SSSAQPMPFFGEILGAFKNL-------------KRPMWILLLVTCLNWIAWFPFLLFDT 318

Qy     294 DFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRA 353
           |::|  :| |   :    |  : || ||| |:||| |   :    || :: | :  |
Db     319 DWMGREVYGG--DSSGSAEQLKLYDRGVRAGALGLMLNSVVLGFTSLGVEVLARGVG-GV 375

Qy     354 VYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPK 413
             |  :  | :   | ||: :| :||  |  |                      :
Db     376 KRLWGIVNFVL---AVCLAMTV-LVTKQAEST-------------------------RR 405

Qy     414 YRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVV 473
           :   :|||          :|                    ||||    |  |  ::  |
Db     406 FATVSGGAK---------VP-------------------LPPP---SGVKAGALALFAV 433

Qy     474 VGEPTEARV-VP----------GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL--SQS 519
           :| |      :|          | |  | | :|: : :: |:  |:  |    |    :
Db     434 MGVPQAITYSIPFALASIFSNTSGAGQGLSLGVLNLSIVIPQMIVSVAAGPWDALFGGGN 493

Qy     520 VTAYMVSAAGLGLVAIYFATQVVFDKSDLAKYSA 553
           : |::| |       |:  | :   : |:    |
Db     494 LPAFVVGAVAALASGIFALTMLPSPQPDMPSAKA 527


RESULT  10
T14339
sucrose-proton transport protein - carrot
N;Alternate names: sucrose/H+ symporter protein
C;Species: Daucus carota (carrot)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T14339
R;Shakya, R.; Sturm, A.
Plant Physiol. 118, 1473-1480, 1998
A;Title: Characterization of source- and sink-specific sucrose/H+ symporters from
A;Reference number: Z17991; MUID:99063785
A;Accession: T14339
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-501 <SHA>
A;Cross-references: EMBL:Y16766; NID:g2969886; PIDN:CAA76367.1; PID:g2969887
A;Experimental source: cultivar Namtaise: leaf
C;Genetics:
A;Gene: SUT1a
C;Superfamily: common tobacco sucrose transport protein


  Query Match             10.7%;  Score 305.5;  DB 2;  Length 501;
  Best Local Similarity   21.6%;  Pred. No. 1.3e-15;
  Matches  124;  Conservative   91;  Mismatches  213;  Indels  145;  Gaps   14;

Qy      18 LLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASDH 77
           || |  :  |::   |  :: : | : |:|:   : :::   ||: ||:  |::|  ||
Db      30 LLRVASVACGIQFGWALQLSLLTPYVQELGIPHAWSSIIWLCGPLSGLLVQPIVGHMSDQ 89

Qy      78 WRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLL--CPDPRPLELALLILGVGLLDFC 135
             :|||||||| |    |:|:: :|  :  : |||    | : : :   ::|  :||
Db      90 CTSKYGRRRPFIVAGGTAIILAVIIIAHSADIGGLLGDTADNKTMAIVAFVIGFWILDVA 149

Qy     136 GQVCFTPLEALLSDLF-RDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLGT 194
             :  | |||:||   |    | | : ::  :::|  |||          |   | |
Db     150 NNMTQGPCRALLADLTGNDARRTRVANAYFSLFMAIGNVLGY----------ATGAYSGW 199

Qy     195 QEECLFGLLTLIFLTC------------VAATLLVAEEAAL---------GPTEPAEGL 232
           :  | | :   : |              :  |  ::  ||          ||    :|
Db     200 YKVFPFSLTSSCTINCANLKSAFYIDIIFIIITTYISISAAKERPRISSQDGPQFSEDGT 259

Qy     233 SAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFY 292
           :   | |              | |  |     :| ::  :  :   :|:    | ||
Db     260 AQ---SGHI-----------EEAFLWELFGTFRLLPGSVWVILLVTCLNWIGWFPFILFD 305
```

```
Qy    293 TDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTR 352
          ||::|  :| | |        : | :|||||: || :   :  : |::|::| : :|:
Db    306 TDWMGREIYGGEP------NQGQSYSDGVRMGAFGLMMNSVVLGITSVLMEKLCRIWGSG 359

Qy    353 AVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLP 412
           ::                 |  |    | |: :: :   :: :
Db    360 FMW----------------------GLSNILMTICFFAMLLITFIAKNMDY-------- 388

Qy    413 KYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRV 472
                                                |  |||    |     : |
Db    389 ------------------------------------GTNPPPN---GIVISALIVFA 406

Qy    473 VVGEPTEARV-VP----------GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL---- 516
          ::| |      ||           | |  | : :|: | :: ||  ||  |   ||
Db    407 ILGIPLAITYSVPYALVSTRIESLGLGQGLSMGVLNLAIVVPQVIVSLGSGPWDQLFGGG 466

Qy    517 -SQSVTAYMVSAAGLGLVAIYFATQVVFDKSDL 548
           | :    :||  ||:|:    :   ||| |
Db    467 NSPAFVVAALSAFAAGLIALIAIRRPRVDKSRL 499


RESULT  11
S48788
sucrose transport protein - tomato (fragment)
C;Species: Lycopersicon esculentum (tomato)
C;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 09-Jun-2000
C;Accession: S48788
R;Buerkle, X.Y.Z.; Frommer, W.B.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48787
A;Accession: S48788
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-428 <BUE>
A;Cross-references: EMBL:X82275
C;Superfamily: common tobacco sucrose transport protein


  Query Match             10.4%;  Score 297.5;  DB 2;  Length 428;
  Best Local Similarity   25.8%;  Pred. No. 4.6e-15;
  Matches  114;  Conservative   82;  Mismatches  163;  Indels   83;  Gaps   18;

Qy      4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
          :||          ::::|  :  |::   |  :: : | :  :|:  :| : :   ||:
Db     25 KLW----------KIIVVASIAAGVQFGWALQLSLLTPYVQLLGIPHRFASFIWLCGPIS 74

Qy     64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDP-----R 118
          |::  |::|  ||:   |:|||||||| | :  : :::|||  |  | |    ||      :
Db     75 GMIVQPVVGYYSDNCSSRFGRRRPFIAAGAALVTIAVFLIGFAADL-GHASGDPLGKGSK 133

Qy    119 PLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPD-HCRQAYSVYAFMISLGGCLGY- 176
          |  :|: ::|  :||    :   |  |||:||        | | : ::|  ::|  |||
Db    134 PRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKSGKMRTANAFFSFFKAVGNILGYA 193

Qy    177 ------LLPAIDWD-TSALAPYLGTQEECLF-GLLTLIFLTCVAATLL----VAEEAALG 224
                |     :  | |   |    : | |  :  |: || :| ||:    : |:  |
Db    194 AGSYSRLFKVFPFSKTKACDMYRANLKSCFFIAIFLLLSLTTLALTLVRENELPEKEELE 253

Qy    225 PTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMA 284
            |   |              :::: |   | :   |  |    || :  | :    :|:|
Db    254 IDEKLSGAG----------KSKVPF--FGEIFGALKDL----PRPMWILLLVTCLNWIA 296

Qy    285 LMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDR 344
             | |: ||:: : :: |    : |    : || ||| |:||| ||  :    || ::
Db    297 WFPFFLYDTDWMAKEVFGG----QVGD--AKLYDLGVRAGALGLLLQSVVLGFMSLGVEF 350

Qy    345 LVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVV---------TASAALTGFT------- 388
          | :: |  |  |  :  | :   | ||: :: |           |:  | | |
Db    351 LGKKIG-GAKRLWGILNFVL---AICLAMTILVTKMAEKSRGHDAAGTLMGPTPGVKIGA 406

Qy    389 ---FSALQI-------LPYTLA 400
             |:|| |       :|: ||
Db    407 LLLFAALGIPLAVTFSIPFALA 428


RESULT  12
S38657
sucrose transport protein ptp1 - common plantain
N;Alternate names: sucrose transporter ptp1
C;Species: Plantago major (common plantain)
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 26-May-2000
C;Accession: S38657
R;Gahrtz, M.; Stolz, J.; Sauer, N.
submitted to the EMBL Data Library, November 1993
A;Reference number: S38657
A;Accession: S38657
A;Molecule type: mRNA
A;Residues: 1-510 <GAH>
A;Cross-references: EMBL:X75764; NID:g415987; PIDN:CAA53390.1; PID:g415988
C;Genetics:
A;Gene: ptp1
C;Superfamily: common tobacco sucrose transport protein
C;Keywords: sugar transport


  Query Match             10.3%;  Score 294;  DB 2;  Length 510;
  Best Local Similarity   24.7%;  Pred. No. 1e-14;
  Matches  126;  Conservative   77;  Mismatches  199;  Indels  108;  Gaps   19;

Qy     18 LLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASDH 77
          : ||  :  |::   |  :: : | :  :|:  |: : :   ||: |::  |::|  ||:
Db     29 IFLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWASYIWLCGPISGMIVQPVVGYFSDN 88

Qy     78 WRGRYGRRRPFIWALSLGIL-LSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
             |:|||||| |   |:: ::: ||   |: | |        |    :|  : : : |
Db     89 CTSRFGRRRPFI-AAGAGLVGVAVVLI---GFAADLGHAGGDSLGDGLKPRAIGVFVFGF 144

Qy    130 GLLDFCGQVCFTPLEALLSDLF-RDPDHCRQAYSVYAFMISLGGCLGYLLPAID------ 182
             :||    :   |  |||:||   :       | | ::| :::|  |||   :
Db    145 WILDVANNMLQGPCRALLADLSGGNTKKMANANSFFSFFMAVGNVLGYAAGSYSRMYKVF 204

Qy    183 --WDTSALAPYLGTQEEC-LFGLLTLIFLTCVAATLL-----VAEEAALGPTEPAEGLSA 234
               | |    |    : | :  :  || || :| :::     ||||     |   :|
Db    205 PFSKTKACDIYCANLKSCFIISITLLITLTTLALSIVREKRHVAEEQV---TAAKKGFKI 261

Qy    235 PSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTD 294
          |                      : | |      :|| :  | :    :|:|   | || ||
Db    262 P--------------------VFPELFGALKDLPRPMWVLLLVTALNWIAWFGFLLFDTD 301

Qy    295 FVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG---- 350
          ::|  :| | :      |    |::|| |:||| |   :     || : : :  |
Db    302 WMGREVY-GETQQHKAPELAVIYNKGVSAGALGLMLNSIVLGFASLGVQYMARALGGVKR 360

Qy    351 -------TRAVYLA-SVAAFPVAAGATCLSHSVAVVTASAALTG--FTFSALQI------ 394
                  |: |  ::    ||:     |: |     |:   |    ||| |
Db    361 LWGVVNFILAICLCMTIVITKVASHHRPYSNGVLQTPESSVKIGALVVFSALGIPLAITF 420

Qy    395 -LPYTLASLYHREKQVFLPKYRGDTGGASSEDSL-----------MTSFLPGPKPG--- 438
           :|: |||:|              | |:    ||             : |   ||
Db    421 SVPFALASIY-----------STTTGSGQGLSLGVLNLAIVIPQMIVSVASGPWDAMFG 468

Qy    439 ----APFPNGHVGAGGSG-----LLPPPPA 459
                |  | | |  ||      :|| |||
Db    469 GGNLPAFVVGAVAAAASGIFAFTMLPSPPA 498


RESULT  13
S51114
sucrose-proton symporter - beet
C;Species: Beta vulgaris (beet)
C;Date: 07-May-1995 #sequence_revision 01-Sep-1995 #text_change 26-May-2000
C;Accession: S51114
R;Westram, A.; Eckhardt, U.; Frommer, W.B.; Riesmeier, J.W.
submitted to the EMBL Data Library, January 1995
A;Description: Sequence of a sugar beet sucrose transporter cDNA.
A;Reference number: S51114
```

A;Accession: S51114
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-523 <WES>
A;Cross-references: EMBL:X83850; NID:g633171; PID:g633172
C;Superfamily: common tobacco sucrose transport protein

Query Match 10.3%; Score 294; DB 2; Length 523;
Best Local Similarity 24.3%; Pred. No. 1e-14;
Matches 125; Conservative 78; Mismatches 199; Indels 112; Gaps 20;

```
Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :| ||  :  |::  |  :: : | :  :|:   :   :   ||: |::  | :|  ||
Db      38 KLALVASIAAGVQFGWALQLSLLTPYVQLLGIPHTWAPYIWLCGPISGMIVQPTVGYYSD 97

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPD-----PRPLELALLILGV 129
              ::||| |||   ::|  |  | :   |: |  |     |     |:|  :|: ::|
Db      98 RCTSKFGRRSPFI---AVGATLVGFAVSLIGFAADIGHATGDPNGNVPKPRAIAVFVVGF 154

Qy     130 GLLDFCGQVCFTPLEALLSDLFRDPD-HCRQAYSVYAFMISLGGCLGY---------LLP 179
           :||        |  |||:|:        | | : ::| ::||   ||          : |
Db     155 WILDVANNTLQGPCRALLADMAAGSQAKTRYANAFFSFFMALGNIGGYAGSYGRLYTVFP 214

Qy     180 AIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSP 239
             | |   |    : | |  :||: :  : |  :| |     |    |     :|
Db     215 FT--HTKACDTYCANLKSCFFISITLLIVLTILALSVVRER----PFTLDEIQEEENLKN 268

Qy     240 HCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEG 299
           :  | ||| |   | |   |  |    |: :  | :    :|:|   | || ||::|:
Db     269 NTGGC-ARLPF--FGQLFGALKDL----PKPMLILLLVTCLNWIAWFPFLLFDTDWMGKE 321

Qy     300 LYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG--TRAVYLA 357
           :| |           : || ||  |:||| :   :  : || :::| :  |   |    :
Db     322 VYGGT------VGEGKAYDMGVHAGALGLMINSVVLGIMSLGIEKLARLVGGVKRLWGIV 375

Qy     358 SVAAFPVAAGATCLSHSVAVVTASA------------------------------ALTGF 387
           ::        | ||: :: :|| ||                              |: |
Db     376 NLIL------AVCLAMTI-LVTKSAEHYRATHHVPGAIGPPLPPPGVKGGALAIFAVLGI 428

Qy     388 TFSALQILPYTLASLYHREK-----------------QVFLPKYRGDTGGASSEDSLMT 429
              :|: |||::                        |:|:    |       |:|
Db     429 PLAITFSIPFALASIFSASSGSGQGLSLGVLNLAIVVPQMFVSVTSGPW------DAL-- 480

Qy     430 SFLPGPKPGAPFPNGHVGAGGSG-----LLPPPP 458
            |  |  |   |  | | |  |      ||||||
Db     481 -FGGGNLPA--FVVGAVAATASAILSFTLLPPPP 511
```

RESULT 14
F96741
probable sucrose transport protein F17M19.4 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: F96741
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: F96741
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-512 <STO>
A;Cross-references: GB:AE005173; NID:g6978914; PIDN:AAF34306.1; GSPDB:GN00141
C;Genetics:
A;Gene: F17M19.4
A;Map position: 1
C;Superfamily: common tobacco sucrose transport protein

Query Match 10.2%; Score 292.5; DB 2; Length 512;
Best Local Similarity 23.8%; Pred. No. 1.3e-14;
Matches 120; Conservative 87; Mismatches 211; Indels 87; Gaps 18

```
Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           ::: |  :  |::   |  :: : | :  :|:  |: : :   ||: |::  |::|  ||
Db      33 KIISVASIAAGVQFGWALQLSLLTPYIQLLGIPHKWSSYMWLCGPISGMIVQPIVGYHSD 92

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
              |:|||||| |    : :|:|||   |: | :         |    |   : : : |
Db      93 RCESRFGRRRPFIAAGVALVAVSVFLI---GFAADMGHSFGDKLENKVRTRAIIIFLTGF 149

Qy     130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LL 178
             ||        |  | |:||   |    | | : ::| :::|  |||          :
Db     150 WFLDVANNTLQGPCRAFLADLAAGDAKKTRVANACFSFFMAVGNVLGYAAGSYTNLHKMF 209

Qy     179 PAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLS 238
           |     | |   |    : | |  :||: :   ::   | ::       | :|      |
Db     210 PFT--MTKACDIYCANLKTCFFLSITLLLIVTFSSLWYVKDK----QWSPPQGDKEEKTS 263

Qy     239 PHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGE 298
              | |   | :   :      | | :  | :  : :|:|   | |: ||::|
Db     264 --------SLFF--FGEIFGAVR----HMKRPMVMLLIVTVINWIAWFPFILYDTDWMGR 309

Qy     299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
            :| |   ::    ::: ||:||: |:|||     :   || :: : :: |  |  |
Db     310 EVYGG--NSDGDERSKKLYDQGVQAGALGLMFNSILLGFVSLGVESIGRKMG-GAKRLWG 366

Qy     359 VAAFPVAAGATCLSHSVAVVTA--------------SAALTGFTFSALQIL--------- 395
           | :| |   |: :| |  :              |: :    ||   :|
Db     367 CVNFILAIG---LAMTVLVTKSAEHHREIAGPLAGPSSGIKAGVFSLFTVLGIPLAITYS 423

Qy     396 -PYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKP----GAPFPN---GHVG 447
            |: |||::            |    |     :: ||  ||       |   |:   | :
Db     424 IPFALASIFSTNSGAGQGLSLGVLNIAICIPQMIVSFSSGPLDAQFGGGNLPSFVVGAIA 483

Qy     448 AGGSGLL-------PPP--PALCGA 463
           |  ||:|        |||  ||: ||
Db     484 AAVSGVLALTVLPSPPPDAPAMSGA 508
```

RESULT 15
T02982
probable sucrose transport protein - rice
C;Species: Oryza sativa (rice)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C;Accession: T02982
R;Hirose, T.; Imaizumi, N.; Scofield, G.N.; Furbank, R.T.; Ohsugi, R.
Plant Cell Physiol. 38, 1389-1396, 1997
A;Title: cDNA cloning and tissue specific expression of a gene for sucrose transp
A;Reference number: Z14809; MUID:98182940
A;Accession: T02982
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-537 <HIR>
A;Cross-references: EMBL:D87819; NID:g2723470; PIDN:BAA24071.1; PID:g2723471
A;Experimental source: cultivar Nipponbare, leaf
C;Genetics:
A;Gene: SUT1
C;Superfamily: common tobacco sucrose transport protein

Query Match 10.2%; Score 292; DB 2; Length 537;
Best Local Similarity 26.7%; Pred. No. 1.5e-14;
Matches 115; Conservative 72; Mismatches 167; Indels 76; Gaps 19

```
Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|:|  ::  |::   |  :: : | :  :|:     :  :   ||: |:|  | :|  ||
Db      51 RLILSGMVAGGVQYGWALQLSLLTPYVQTLGLSHALTSFMWLCGPIAGMVVQPCVGLYSD 110
```

```
Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLI---PRAGWLAGLLCPD------PRPLELALLIL 127
           ::|||||:|    : | |:: :|       |:  |    |        |     : :|
Db     111 RCTSKWGRRRPYILTGCVLICLAVVVIGFSADIGYAMGDTKEDCSVYHGSRWHAAIVYVL 170

Qy     128 GVGLLDFCGQVCFTPLEALLSDLF--RDPDHCRQAYSVYAFMISLGGCLGYLLPAID--- 182
           |  ||||       |  ||::||     |     | |::   :::|  |||   : :
Db     171 GFWLLDFSNNTVQGPARALMADLSGRHGPG---TANSIFCSWMAMGNILGYSSGSTNNWH 227

Qy     183 -WDTSALAPYLGTQEEC-----LFG--LLTLIFLT-CVAATLLVAEE------AALGPT- 226
            |      |:| |:  |     | |  |: :|||: |:  ||: |:|      ||| ||
Db     228 KW-----FPFLKTRACCEACANLKGAFLVAVIFLSLCLVITLIFAKEVPFKGNAAL-PTK 281

Qy     227 --EPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMA 284
             ||||           |      ||||   :| :              | |  | :|::
Db     282 SNEPAEPEGTG-------PLAVLKGFRNLPTGMPSV-------------LIVTGL-TWLS 320

Qy     285 LMTFTLFYTDFVGEGLYQGVPR-AEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMD 343
             | |: ||::|  :| | |:  :|  ||   :::||| |: || |   :    | :::
Db     321 WFPFILYDTDWMGREIYHGDPKGTDPQIEA---FNQGVRAGAFGLLLNSIVLGFSSFLIE 377

Qy     344 RLVQRFGTRAVYLASVAAFPVAAGATCL--------SHSVA--VVTASAALTGFTFSALQ 393
           : :: | | |:: |     :|  || |          |      :||  ::
Db     378 PMCRKVGPRVVWVTSNFLVCIAMAATALISFWSLKDFHGTVQKAITADKSIKAVCLVLFA 437

Qy     394 ILPYTLASLY 403
           |   || ||
Db     438 FLGVPLAVLY 447
```

Search completed: February 26, 2002, 07:57:01
Job time: 131 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 07:56:15 ; Search time 14.08 Seconds
(without alignments)
883.830 Million cell updates/sec

Title: US-09-605-783A-113
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2861	100.0	553	4	US-09-020-956-113	Sequence 113, App
2	2861	100.0	553	4	US-09-030-607-113	Sequence 113, App
3	1287	45.0	255	4	US-09-071-710-36	Sequence 36, Appl
4	1287	45.0	255	4	US-09-525-397-36	Sequence 36, Appl
5	330	11.5	516	1	US-08-356-340-4	Sequence 4, Appli
6	330	11.5	516	2	US-08-786-555-4	Sequence 4, Appli
7	324.5	11.3	525	1	US-08-356-340-2	Sequence 2, Appli
8	324.5	11.3	525	2	US-08-786-555-2	Sequence 2, Appli
9	243	8.5	44	4	US-09-071-710-37	Sequence 37, Appl
10	243	8.5	44	4	US-09-525-397-37	Sequence 37, Appl
11	146	5.1	27	4	US-09-071-710-39	Sequence 39, Appl
12	146	5.1	27	4	US-09-525-397-39	Sequence 39, Appl
13	135	4.7	27	4	US-09-071-710-38	Sequence 38, Appl
14	135	4.7	27	4	US-09-525-397-38	Sequence 38, Appl
15	109.5	3.8	455	1	US-08-035-928-2	Sequence 2, Appli
16	106.5	3.7	436	6	5432081-10	Patent No. 5432081
17	106	3.7	427	4	US-09-199-737-4	Sequence 4, Appli
18	98	3.4	713	3	US-09-335-409-11	Sequence 11, Appl
19	97.5	3.4	3011	3	US-08-811-566-20	Sequence 20, Appl
20	97	3.4	370	4	US-09-199-737-2	Sequence 2, Appli
21	95.5	3.3	528	2	US-08-403-852D-21	Sequence 21, Appl
22	95.5	3.3	528	3	US-08-510-646B-22	Sequence 22, Appl
23	95.5	3.3	528	4	US-09-231-818-21	Sequence 21, Appl
24	95	3.3	907	3	US-08-863-102-2	Sequence 2, Appli
25	93.5	3.3	435	6	5268463-9	Patent No. 5268463
26	93.5	3.3	2995	4	US-08-444-818-138	Sequence 138, App
27	93.5	3.3	3011	1	US-08-440-103-36	Sequence 36, Appl
28	93.5	3.3	3011	1	US-08-440-542-36	Sequence 36, Appl
29	93.5	3.3	3011	1	US-07-910-760-10	Sequence 10, Appl
30	93.5	3.3	3011	1	US-08-440-519-10	Sequence 10, Appl
31	93.5	3.3	3011	1	US-08-231-368-36	Sequence 36, Appl
32	93.5	3.3	3011	1	US-08-440-210-36	Sequence 36, Appl
33	93.5	3.3	3011	4	US-09-388-874-2	Sequence 2, Appli
34	93.5	3.3	3011	4	US-09-046-604-36	Sequence 36, Appl
35	92.5	3.2	649	4	US-08-462-467B-20	Sequence 20, Appl
36	92.5	3.2	1495	4	US-08-462-467B-12	Sequence 12, Appl
37	90.5	3.2	967	1	US-08-188-281B-13	Sequence 13, Appl
38	90.5	3.2	967	5	PCT-US94-07280-13	Sequence 13, Appl
39	90.5	3.2	967	5	PCT-US95-01087-13	Sequence 13, Appl
40	90.5	3.2	1648	1	US-08-188-281B-12	Sequence 12, Appl
41	90.5	3.2	1648	5	PCT-US94-07280-12	Sequence 12, Appl
42	90.5	3.2	1648	5	PCT-US95-01087-12	Sequence 12, Appl
43	90.5	3.2	3011	1	US-08-188-281B-1	Sequence 1, Appli
44	90.5	3.2	3011	1	US-08-453-552-1	Sequence 1, Appli
45	90.5	3.2	3011	2	US-08-710-637-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-09-020-956-113
; Sequence 113, Application US/09020956
; Patent No. 6261562
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillin, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHO
;    NUMBER OF SEQUENCES:  178
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/020,956
;      FILING DATE:  09-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  113:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  553 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-020-956-113


  Query Match           100.0%;  Score 2861;  DB 4;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 3.9e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy      61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy     121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy     181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy     301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy     361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy     421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy     481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy     541 VVFDKSDLAKYSA 553
           |||||||||||||
Db     541 VVFDKSDLAKYSA 553


RESULT   2
US-09-030-607-113
; Sequence 113, Application US/09030607
; Patent No. 6262245
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
;    NUMBER OF SEQUENCES:  224
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,607
;      FILING DATE:  25-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  113:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  553 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-030-607-113


  Query Match             100.0%;  Score 2861;  DB 4;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy      61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy     121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy     181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy     301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy     361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy     421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy     481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy     541 VVFDKSDLAKYSA 553
           |||||||||||||
Db     541 VVFDKSDLAKYSA 553


RESULT   3
US-09-071-710-36
; Sequence 36, Application US/09071710
; Patent No. 6130043
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
```

```
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,710
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/850,713
;      FILING DATE:  02-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  255 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6130043e
US-09-071-710-36


  Query Match             45.0%;  Score 1287;  DB 4;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-118;
  Matches  255;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 60

Qy    359 VAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDT 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 VAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDT 120

Qy    419 GGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPT 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPT 180

Qy    479 EARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFA 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 EARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFA 240

Qy    539 TQVVFDKSDLAKYSA 553
          |||||||||||||||
Db    241 TQVVFDKSDLAKYSA 255


RESULT   4
US-09-525-397-36
; Sequence 36, Application US/09525397
; Patent No. 6252047
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/525,397
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/071,710
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  255 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6252047e
US-09-525-397-36


  Query Match             45.0%;  Score 1287;  DB 4;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-118;
  Matches  255;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 60

Qy    359 VAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDT 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 VAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDT 120

Qy    419 GGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPT 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPT 180

Qy    479 EARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFA 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 EARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFA 240

Qy    539 TQVVFDKSDLAKYSA 553
          |||||||||||||||
Db    241 TQVVFDKSDLAKYSA 255
```

```
RESULT   5
US-08-356-340-4
; Sequence 4, Application US/08356340
; Patent No. 5608146
;  GENERAL INFORMATION:
;    APPLICANT:  FROMMER, Wolf-Bernd
;    APPLICANT:  RIESMEIER, Jorg
;    TITLE OF INVENTION:  DNA SEQUENCES WITH OLIGOSACCHARIDE
;    TITLE OF INVENTION:  TRANSPORTER, PLASMIDS, BACTERIA AND PLANTS CONTAINING A
;    TITLE OF INVENTION:  TRANSPORTER AS WELL AS A PROCESS FOR THE PREPARATION AND
;    TITLE OF INVENTION:  TRANSFORMATION OF YEAST STRAINS FOR THE IDENTIFI...
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ostrolenk, Faber, Gerb & Soffen
;      STREET:  1180 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  US
;      ZIP:  10036-8403
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,340
;      FILING DATE:  21-DEC-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/EP93/01604
;      FILING DATE:  22-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P42 20 759.2
;      FILING DATE:  24-JUN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meilman, Edward A.
;      REGISTRATION NUMBER:  24,735
;      REFERENCE/DOCKET NUMBER:  P/951-106
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 382-0700
;      TELEFAX:  (212) 382-0888
;      TELEX:  236925
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  516 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-356-340-4


  Query Match             11.5%;  Score 330;  DB 1;  Length 516;
  Best Local Similarity   24.7%;  Pred. No. 6.9e-24;
  Matches  132;  Conservative   93;  Mismatches  197;  Indels  112;  Gaps   21;

Qy       4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
           :||          ::::|  :  |::   |  :: : | :  :|:  || : :   ||:
Db      29 KLW----------KIIVVASIAAGVQFGWALQLSLLTPYVQLLGIPHKFASFIWLCGPIS 78

Qy      64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPD 116
           |::  |::|  ||:   |:|||||||| | :  :::::|||   |: | |       |
Db      79 GMIVQPVVGYYSDNCSSRFGRRRPFIAAGAALVMIAVFLI---GFAADLGHASGDTLGKG 135

Qy     117 PRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPD-HCRQAYSVYAFMISLGGCLG 175
            :|  :|: ::|  :||    :   |  |||:||        | | : ::| :::|  ||
Db     136 FKPRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKSGRMRTANAFFSFFMAVGNILG 195

Qy     176 YLLPAIDW--------DTSALAPYLGTQEECLF-GLLTLIFLTCVAATLLVAEE-AALGP 225
           |   :            | |   |    : | |  :  |: || :| ||:   |
Db     196 YAAGSYSHLFKVFPFSKTKACDMYCANLKSCFFIAIFLLLSLTTIALTLVRENELPEKDE 255

Qy     226 TEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMAL 285
           |  | |:          :::: |   | :   | :|    || :  | :    :|:|
Db     256 QEIDEKLAGAG--------KSKVPF--FGEIFGALKEL----PRPMWILLLVTCLNWIAW 301

Qy     286 MTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRL 345
           | |: ||:: : :: |    : |    | || ||| |::|| ||  :    || :: |
Db     302 FPFFLYDTDWMAKEVFGG----QVGD--ARLYDLGVRAGAMGLLLQSVVLGFMSLGVEFL 355

Qy     346 VQRFGTRAVYLASVAAFPVAAGATCLSHSVAVV---------TASAALTGFT-------- 388
           :: |  |  |  :  | :   | ||: :: |             :  | | |
Db     356 GKKIG-GAKRLWGILNFVL---AICLAMTILVTKMAEKSRQHDPAGTLMGPTPGVKIGAL 411

Qy     389 --FSALQI-------LPYTLASLYHREK------------QVFLPKYRGDTGGASSEDS 426
             |:|| |        :|: |||::   :              : :|:      |   :|
Db     412 LLFAALGIPLAATFSIPFALASIFSSNRGSGQGLSLGVLNLAIVVPQMLVSLVGGPWDDL 471

Qy     427 LMTSFLPGPKPGAPFPNGHVGAGGSG-----LLPPPPALCGASACDVSVRVVVG 475
               |||       |  | | |  |       :|| |||       |    | :|
Db     472 FGGGNLPG------FVVGAVAAAASAVLALTMLPSPPA-------DAKPAVAMG 512


RESULT   6
US-08-786-555-4
; Sequence 4, Application US/08786555B
; Patent No. 5981181
; GENERAL INFORMATION:
;  APPLICANT: FROMMER, Wolf-Bernd
;  APPLICANT: RIESMEIER, Jorg
;  TITLE OF INVENTION: DNA SEQUENCES WITH OLIGOSACCHARIDE TRANSPORTER,
;  TITLE OF INVENTION: PLASMIDS, BACTERIA AND PLANTS CONTAINING A TRANSPORTER
;  TITLE OF INVENTION: AS WELL AS A PROCESS FOR PREPARATION AND TRANSFORMATION
;  TITLE OF INVENTION: OF YEAST STRAINS FOR THE IDENTIFICATION
;  FILE REFERENCE: Frommer
;  CURRENT APPLICATION NUMBER: US/08/786,555B
;  CURRENT FILING DATE: 1997-01-21
;  EARLIER APPLICATION NUMBER: 08/356,340
;  EARLIER FILING DATE: 1994-12-21
;  EARLIER APPLICATION NUMBER: PCT/EP93/01604
;  EARLIER FILING DATE: 1993-06-22
;  EARLIER APPLICATION NUMBER: DE P4220759.2
;  EARLIER FILING DATE: 1992-06-24
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 516
;   TYPE: PRT
;   ORGANISM: Spinacia oleracea
US-08-786-555-4


  Query Match             11.5%;  Score 330;  DB 2;  Length 516;
  Best Local Similarity   24.7%;  Pred. No. 6.9e-24;
  Matches  132;  Conservative   93;  Mismatches  197;  Indels  112;  Gaps   21;

Qy       4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
           :||          ::::|  :  |::   |  :: : | :  :|:  || : :   ||:
Db      29 KLW----------KIIVVASIAAGVQFGWALQLSLLTPYVQLLGIPHKFASFIWLCGPIS 78

Qy      64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPD 116
           |::  |::|  ||:   |:|||||||| | :  :::::|||   |: | |       |
Db      79 GMIVQPVVGYYSDNCSSRFGRRRPFIAAGAALVMIAVFLI---GFAADLGHASGDTLGKG 135

Qy     117 PRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPD-HCRQAYSVYAFMISLGGCLG 175
            :|  :|: ::|  :||    :   |  |||:||        | | : ::| :::|  ||
Db     136 FKPRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKSGRMRTANAFFSFFMAVGNILG 195

Qy     176 YLLPAIDW--------DTSALAPYLGTQEECLF-GLLTLIFLTCVAATLLVAEE-AALGP 225
           |   :            | |   |    : | |  :  |: || :| ||:   |
Db     196 YAAGSYSHLFKVFPFSKTKACDMYCANLKSCFFIAIFLLLSLTTIALTLVRENELPEKDE 255

Qy     226 TEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMAL 285
           |  | |:          :::: |   | :   | :|    || :  | :    :|:|
Db     256 QEIDEKLAGAG--------KSKVPF--FGEIFGALKEL----PRPMWILLLVTCLNWIAW 301

Qy     286 MTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRL 345
```

```
Db    302 FPFFLYDTDWMAKEVFGG----QVGD--ARLYDLGVRAGAMGLLLQSVVLGFMSLGVEFL 355

Qy    346 VQRFGTRAVYLASVAAFPVAAGATCLSHSVAVV---------TASAALTGFT-------- 388

Db    356 GKKIG-GAKRLWGILNFVL---AICLAMTILVTKMAEKSRQHDPAGTLMGPTPGVKIGAL 411

Qy    389 --FSALQI-------LPYTLASLYHREK------------QVFLPKYRGDTGGASSEDS 426

Db    412 LLFAALGIPLAATFSIPFALASIFSSNRGSGQGLSLGVLNLAIVVPQMLVSLVGGPWDDL 471

Qy    427 LMTSFLPGPKPGAPFPNGHVGAGGSG-----LLPPPPALCGASACDVSVRVVVG 475

Db    472 FGGGNLPG------FVVGAVAAAASAVLALTMLPSPPA-------DAKPAVAMG 512


RESULT   7
US-08-356-340-2
; Sequence 2, Application US/08356340
; Patent No. 5608146
;  GENERAL INFORMATION:
;    APPLICANT:  FROMMER, Wolf-Bernd
;    APPLICANT:  RIESMEIER, Jorg
;    TITLE OF INVENTION:  DNA SEQUENCES WITH OLIGOSACCHARIDE
;    TITLE OF INVENTION:  TRANSPORTER, PLASMIDS, BACTERIA AND PLANTS CONTAINING A
;    TITLE OF INVENTION:  TRANSPORTER AS WELL AS A PROCESS FOR THE PREPARATION AND
;    TITLE OF INVENTION:  TRANSFORMATION OF YEAST STRAINS FOR THE IDENTIFI...
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ostrolenk, Faber, Gerb & Soffen
;      STREET:  1180 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  US
;      ZIP:  10036-8403
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,340
;      FILING DATE:  21-DEC-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/EP93/01604
;      FILING DATE:  22-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P42 20 759.2
;      FILING DATE:  24-JUN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meilman, Edward A.
;      REGISTRATION NUMBER:  24,735
;      REFERENCE/DOCKET NUMBER:  P/951-106
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 382-0700
;      TELEFAX:  (212) 382-0888
;      TELEX:  236925
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  525 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-356-340-2


  Query Match           11.3%;  Score 324.5;  DB 1;  Length 525;
  Best Local Similarity   23.2%;  Pred. No. 2.4e-23;
  Matches  129;  Conservative   91;  Mismatches  196;  Indels  141;  Gaps   20;

Qy     23 LLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSA 74

Db     35 LKKLGLVASVAAGVQFGWALQLSLLTPYVQLLGIPHTWAAYIWLCGPISGMIVQPLVGYY 94

Qy     75 SDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPDP-----RPLELALLIL 127

Db     95 SDRCTSRFGRRRPFIAAGAALVAVAVGLI---GFAADIGAASGDPTGNVAKPRAIAVFVV 151

Qy    128 GVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID---- 182

Db    152 GFWILDVANNTLQGPCRALLADMAAGSQTKTRYANAFFSFFMALGNIGGYAAGSYSRLYT 211

Qy    183 ----WDTSALAPYLGTQEECLFGLLT-LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237

Db    212 VFPFTKTAACDVYCANLKSCFFISITLLIVLTILALSVVKERQITIDEIQEEEDLKNRNN 271

Qy    238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297

Db    272 SSGC----ARLPF--FGQLIGALKDL----PKPMLILLLVTALNWIAWFPFLLFDTDWMG 321

Qy    298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG--TRAVY 355

Db    322 KEVYGGT------VGEGKLYDQGVHAGALGLMINSVVLGVMSLSIEGLARMVGGAKRLWG 375

Qy    356 LASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYR 415

Db    376 IVNIIL------AVCLAMTV-LVTKSAE-------------------HFRDSHHIM---- 405

Qy    416 GDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVG 475

Db    406 ---------------------------------GSAVPPPPPA--GVKGGALAIFAVLG 429

Qy    476 EPTEARV-VP-----------GRGICLDLAILDSAFLLSQV----------------AP 506

Db    430 IPLAITFSIPLALASIFSASSGSGQGLSLGVLNLAIVVPQMFVSVTSGPWDAMFGGGNLP 489

Qy    507 SLFMGSIVQLSQSVTAY 523

Db    490 AFVVGAVAATASAVLSF 506


RESULT   8
US-08-786-555-2
; Sequence 2, Application US/08786555B
; Patent No. 5981181
; GENERAL INFORMATION:
;  APPLICANT: FROMMER, Wolf-Bernd
;  APPLICANT: RIESMEIER, Jorg
;  TITLE OF INVENTION: DNA SEQUENCES WITH OLIGOSACCHARIDE TRANSPORTER,
;  TITLE OF INVENTION: PLASMIDS, BACTERIA AND PLANTS CONTAINING A TRANSPORTER
;  TITLE OF INVENTION: AS WELL AS A PROCESS FOR PREPARATION AND TRANSFORMATION
;  TITLE OF INVENTION: OF YEAST STRAINS FOR THE IDENTIFICATION
;  FILE REFERENCE: Frommer
;  CURRENT APPLICATION NUMBER: US/08/786,555B
;  CURRENT FILING DATE: 1997-01-21
;  EARLIER APPLICATION NUMBER: 08/356,340
;  EARLIER FILING DATE: 1994-12-21
;  EARLIER APPLICATION NUMBER: PCT/EP93/01604
;  EARLIER FILING DATE: 1993-06-22
;  EARLIER APPLICATION NUMBER: DE P4220759.2
;  EARLIER FILING DATE: 1992-06-24
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 525
;   TYPE: PRT
;   ORGANISM: Spinacia oleracea
US-08-786-555-2


  Query Match           11.3%;  Score 324.5;  DB 2;  Length 525;
  Best Local Similarity   23.2%;  Pred. No. 2.4e-23;
  Matches  129;  Conservative   91;  Mismatches  196;  Indels  141;  Gaps   20;

Qy     23 LLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSA 74
```

```
          |   ||   :|||: :        : | :  :|:   :   :   ||: |::  ||:|
Db     35 LKKLGLVASVAAGVQFGWALQLSLLTPYVQLLGIPHTWAAYIWLCGPISGMIVQPLVGYY 94

Qy     75 SDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPDP-----RPLELALLIL 127
          ||    |:|||||| | :  : ::: ||   |: |  |    ||     :|  :|: ::
Db     95 SDRCTSRFGRRRPFIAAGAALVAVAVGLI---GFAADIGAASGDPTGNVAKPRAIAVFVV 151

Qy    128 GVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID---- 182
          |  :||        |  |||:|:        | | : ::| ::||   ||   :
Db    152 GFWILDVANNTLQGPCRALLADMAAGSQTKTRYANAFFSFFMALGNIGGYAAGSYSRLYT 211

Qy    183 ----WDTSALAPYLGTQEECLFGLLT-LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
              |:|   |    : | |  :| || || :| :::   :  :   :  | |   :
Db    212 VFPFTKTAACDVYCANLKSCFFISITLLIVLTILALSVVKERQITIDEIQEEEDLKNRNN 271

Qy    238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
          |  |    ||| |   | |:  |  |    |: :  | :    :|:|   | || ||::|
Db    272 SSGC----ARLPF--FGQLIGALKDL----PKPMLILLLVTALNWIAWFPFLLFDTDWMG 321

Qy    298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG--TRAVY 355
          : :| |          : ||:||  |:||| :   :  | || :: | :  |   |
Db    322 KEVYGGT------VGEGKLYDQGVHAGALGLMINSVVLGVMSLSIEGLARMVGGAKRLWG 375

Qy    356 LASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYR 415
          : ::       | ||: :| :|| ||                  : |:    :
Db    376 IVNIIL------AVCLAMTV-LVTKSAE-----------------HFRDSHHIM---- 405

Qy    416 GDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVG 475
                                          || : |||||  |     :::  |:|
Db    406 ---------------------------------GSAVPPPPPA--GVKGGALAIFAVLG 429

Qy    476 EPTEARV-VP----------GRGICLDLAILDSAFLLSQV----------------AP 506
          |     :|          | |  | | :|: | :: |:                   |
Db    430 IPLAITFSIPLALASIFSASSGSGQGLSLGVLNLAIVVPQMFVSVTSGPWDAMFGGGNLP 489

Qy    507 SLFMGSIVQLSQSVTAY 523
          :  :|::   : :| ::
Db    490 AFVVGAVAATASAVLSF 506


RESULT   9
US-09-071-710-37
; Sequence 37, Application US/09071710
; Patent No. 6130043
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,710
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/850,713
;      FILING DATE:  02-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6130043e
US-09-071-710-37


  Query Match             8.5%;  Score 243;  DB 4;  Length 44;
  Best Local Similarity   100.0%;  Pred. No. 6.4e-17;
  Matches   44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    403 YHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHV 446
          |||||||||||||||||||||||||||||||||||||||||||||
Db      1 YHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHV 44


RESULT  10
US-09-525-397-37
; Sequence 37, Application US/09525397
; Patent No. 6252047
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/525,397
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/071,710
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Becker, Cheryl L.
;       REGISTRATION NUMBER:  35,441
;       REFERENCE/DOCKET NUMBER:  6083.US.P1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  847/935-1729
;       TELEFAX:  847/938-2623
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  37:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  44 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  No. 6252047e
US-09-525-397-37


  Query Match             8.5%;  Score 243;  DB 4;  Length 44;
  Best Local Similarity   100.0%;  Pred. No. 6.4e-17;
  Matches   44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     403 YHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHV 446
           |||||||||||||||||||||||||||||||||||||||||||||
Db       1 YHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHV 44


RESULT  11
US-09-071-710-39
; Sequence 39, Application US/09071710
; Patent No. 6130043
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,710
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/850,713
;      FILING DATE:  02-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6130043e
US-09-071-710-39


  Query Match             5.1%;  Score 146;  DB 4;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-08;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     299 GLYQGVPRAEPGTEARRHYDEGVRMGS 325
           |||||||||||||||||||||||||||
Db       1 GLYQGVPRAEPGTEARRHYDEGVRMGS 27


RESULT  12
US-09-525-397-39
; Sequence 39, Application US/09525397
; Patent No. 6252047
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/525,397
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/071,710
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6252047e
```

```
US-09-525-397-39


  Query Match             5.1%;  Score 146;  DB 4;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-08;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     299 GLYQGVPRAEPGTEARRHYDEGVRMGS 325
           ||||||||||||||||||||||||||||
Db       1 GLYQGVPRAEPGTEARRHYDEGVRMGS 27


RESULT  13
US-09-071-710-38
; Sequence 38, Application US/09071710
; Patent No. 6130043
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,710
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/850,713
;      FILING DATE:  02-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6130043e
US-09-071-710-38


  Query Match             4.7%;  Score 135;  DB 4;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-06;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     471 RVVVGEPTEARVVPGRGICLDLAILDS 497
           ||||||||||||||||||||||||||||
Db       1 RVVVGEPTEARVVPGRGICLDLAILDS 27


RESULT  14
US-09-525-397-38
; Sequence 38, Application US/09525397
; Patent No. 6252047
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/525,397
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/071,710
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6252047e
US-09-525-397-38


  Query Match             4.7%;  Score 135;  DB 4;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-06;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     471 RVVVGEPTEARVVPGRGICLDLAILDS 497
           ||||||||||||||||||||||||||||
Db       1 RVVVGEPTEARVVPGRGICLDLAILDS 27


RESULT  15
US-08-035-928-2
```

; Sequence 2, Application US/08035928
; Patent No. 5538844
; GENERAL INFORMATION:
; APPLICANT: Duyao, Mabel P.
; APPLICANT: MacDonald, Marcy E.
; APPLICANT: Gusella, James F.
; TITLE OF INVENTION: A No. 5538844el Transport Protein Gene from
; TITLE OF INVENTION: the Huntington's Disease Region
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Avenue N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/035,928
; FILING DATE: 19930323
; CLASSIFICATION: 435
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 466-0800
; TELEFAX: (202) 833-8716
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 455 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-035-928-2

Query Match 3.8%; Score 109.5; DB 1; Length 455;
Best Local Similarity 22.9%; Pred. No. 0.02;
Matches 107; Conservative 55; Mismatches 162; Indels 143; Gaps 22;

```
Qy      19 LLVNLLTFGLEVCLAAGITYV-----PPL-------------LLEVGVEEKFMTMVLG-- 58
           ||::|| | | : |  |:         ||              : : ||::: ::: |
Db      32 LLLDLLAFTLLLPLLPGLLESHGRAHDPLYGSWQGGVDWFATAIGMPVEKRYNSVLFGGL 91

Qy      59 IG---PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCP 115
           ||    ||  :| || |: ||         ||| :    :|:  |  :   :   |  |
Db      92 IGSAFSVLQFLCAPLTGATSDCL-----GRRPVMLLCLMGVATSYAVWATSRSFAAFLA- 145

Qy     116 DPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLG 175
                : || |:      | |  :   |:::||   |    |  :|     |||  ||
Db     146 -------SRLIGGISK----GNVSLS--TAIVADL-GSPLARSQGMAVIGVAFSLGFTLG 191

Qy     176 YLLPA---IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGL 232
            :| |   ::     :||:       ||    |:|: |     |  |:
Db     192 PMLGASLPLE-----MAPWFA----LLFAASDLLFIFCFLPETLPLEK------------ 230

Qy     233 SAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMP-------------RTLRRLFVAE 278
            |||::          | ||:   ||  |  |  |                 :||||
Db     231 RAPSIA---------LGFRDAADLLSPLALL--RFSAVARGQDPPSGDRLSSLRRL---G 276

Qy     279 LCSWMALMTFT--------LFYTDFVGEGLYQG----------------VPRAEPGTEA 313
           |  :: |  |:         | :  |    | ||                |  || |
Db     277 LVYFLYLFLFSGLEYTLSFLTHQRFQFSSLQQGKMFFLIGLTMATIQGAYARRIHPGGEV 336

Qy     314 RR---------------HYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRF------GTR 352
                           :   : :  ||| |    : |    :  :|  :      ||
Db     337 AAVKRALLLLVPAFLLIGWGRSLPVLGLGLLLYSFAAAVVVPCLSSVVAGYGSPGQKGTV 396

Qy     353 AVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTL 399
               | |: |   |||    :    :  | |  |  |:| | :||: |
Db     397 MGTLRSLGALARAAGPLVAASVYWLAGAQACFT--TWSGLFLLPFFL 441
```

Search completed: February 26, 2002, 07:57:56
Job time: 101 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2002, 07:52:50 ; Search time 25.01 Seconds
(without alignments)
1637.846 Million cell updates/sec

Title: US-09-605-783A-113
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2861	100.0	553	19	AAW71869	Amino acid encoded
2	2861	100.0	553	19	AAW69385	Prostate tumour sp
3	2861	100.0	553	21	AAB28527	Protein encoded by
4	2861	100.0	553	21	AAY82002	Human immunogenic
5	2861	100.0	553	22	AAU04961	Human prostate tum
6	2861	100.0	553	22	AAM01117	Human prostate-spe
7	2861	100.0	553	22	AAG99002	Human prostate-spe
8	2861	100.0	553	22	AAG62150	Human P501S invent
9	2861	100.0	553	22	AAB74800	Prostate tumour an
10	2861	100.0	1079	22	AAB74830	Prostate tumour an
11	2596	90.7	595	22	AAM01318	Alpha prepro-P501S
12	1417.5	49.5	371	22	AAM01230	P553S splice varia
13	1416	49.5	371	22	AAE01362	Human gene 11 enco
14	1403.5	49.1	400	22	AAM01262	Ra12-P501S-E2 cons
15	1287	45.0	255	20	AAW85068	Protein encoded by
16	1287	45.0	255	21	AAB29268	Human prostate-rel
17	1287	45.0	255	22	AAU04205	Prostate-specific
18	1150	40.2	231	21	AAY54369	Amino acid sequenc
19	1120	39.1	252	22	AAE01423	Human secreted pro
20	457.5	16.0	748	22	AAM40227	Human polypeptide
21	426	14.9	123	22	AAM01228	P553S splice varia
22	357	12.5	494	20	AAY41119	Soybean sucrose tr
23	337.5	11.8	523	20	AAY41129	Vicia faba sucrose
24	326.5	11.4	667	20	AAY41118	Rice sucrose trans
25	325	11.4	563	20	AAY41125	Wheat sucrose tran
26	323.5	11.3	507	21	AAG30206	Arabidopsis thalia
27	323.5	11.3	512	21	AAG30205	Arabidopsis thalia
28	323.5	11.3	539	21	AAG30204	Arabidopsis thalia
29	313	10.9	497	20	AAY41116	Corn sucrose trans
30	312	10.9	58	22	AAM01188	Human prostate-spe
31	312	10.9	58	22	AAG99073	Human prostate-spe
32	311	10.9	533	20	AAY41128	Ricinus communis s
33	305.5	10.7	501	20	AAY41126	Daucus carota sucr
34	303.5	10.6	519	20	AAY41114	Corn sucrose trans
35	303.5	10.6	522	20	AAY41123	Wheat sucrose tran
36	300	10.5	522	20	AAY41124	Wheat sucrose tran
37	292	10.2	537	20	AAY41127	Oryza sativa sucro
38	291	10.2	619	21	AAB40554	Human ORFX ORF318
39	289	10.1	150	22	AAM01229	P553S splice varia
40	267	9.3	267	22	AAM42013	Human polypeptide
41	243	8.5	44	20	AAW85069	PS108 protein deri
42	243	8.5	44	21	AAB29269	Human prostate-rel
43	243	8.5	44	22	AAU04206	Prostate-specific
44	235.5	8.2	322	20	AAY41120	Soybean sucrose tr
45	197.5	6.9	400	20	AAY41117	Rice sucrose trans

ALIGNMENTS

RESULT 1
AAW71869
ID AAW71869 standard; Protein; 553 AA.
XX
AC AAW71869;
XX
DT 06-JAN-1999 (first entry)
XX
DE Amino acid encoded by prostate tumour clone L1-12.
XX
KW Prostate; cancer; tumour; vaccine; immunogen; clone.
XX
OS Homo sapiens.
XX
PN WO9837093-A2.
XX
PD 27-AUG-1998.
XX
PF 25-FEB-1998; 98WO-US03492.
XX
PR 09-FEB-1998; 98US-0020956.
PR 25-FEB-1997; 97US-0806099.
PR 01-AUG-1997; 97US-0904804.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Xu J;
XX
DR WPI; 1998-609886/51.
DR N-PSDB; AAV61201.
XX
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
XX

PS Example 1; Page 82-84; 130pp; English.
XX
CC The present sequence is an immunogenic portion of a prostate tumour
CC protein. The immunogen, or the DNA encoding it, can be used as a
CC vaccine for the treatment of prostate cancer. The immunogen was
CC isolated from a prostate tumour cDNA library obtained by subtracting
CC a prostate tumour cDNA expression library with a normal tissue cDNA
CC library.
XX
SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 19; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.2e-269;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvpplllevgveekfmtmvlgig 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 vvfdksdlakysa 553
```

RESULT 2
AAW69385
ID AAW69385 standard; Protein; 553 AA.
XX
AC AAW69385;
XX
DT 08-DEC-1998 (first entry)
XX
DE Prostate tumour specific gene clone L1-12 protein.
XX
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy.
XX
OS Homo sapiens.
XX
PN WO9837418-A2.

XX
PD 27-AUG-1998.
XX
PF 25-FEB-1998; 98WO-US03690.
XX
PR 09-FEB-1998; 98US-0904809.
PR 25-FEB-1997; 97US-0806596.
PR 01-AUG-1997; 97US-0904809.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Xu J;
XX
DR WPI; 1998-480805/41.
DR N-PSDB; AAV58586.
XX
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
XX
PS Example 1; Page 87-89; 141pp; English.
XX
CC This sequence is encoded by a human prostate tumour specific gene, and
CC can be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC this protein sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
XX
SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 19; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.2e-269;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvpplllevgveekfmtmvlgig 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540
```

```
Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 vvfdksdlakysa 553


RESULT   3
AAB28527
ID   AAB28527 standard; Protein; 553 AA.
XX
AC   AAB28527;
XX
DT   07-FEB-2001  (first entry)
XX
DE   Protein encoded by human breast tumour cDNA clone P501S.
XX
KW   Human; breast tumour antigen; cytostatic; immunotherapy;
KW   breast cancer; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200061756-A2.
XX
PD   19-OCT-2000.
XX
PF   10-APR-2000; 2000WO-US09688.
XX
PR   09-APR-1999;   99US-0288950.
PR   02-JUL-1999;   99US-0346327.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Reed SG,  Xu J,  Dillon DC;
XX
DR   WPI; 2000-638568/61.
DR   N-PSDB; AAC79473.
XX
PT   A novel isolated polypeptide comprising an immunogenic portion of a
PT   breast cancer protein useful in the detection and treatment of breast
PT   cancer -
XX
PS   Claim 2; Page 92-93; 95pp; English.
XX
CC   The present sequence is encoded by a cDNA sequence which was isolated
CC   from a breast tumour cDNA library. It is provided in a specification
CC   relating to compounds for immunotherapy and diagnosis of breast cancer.
CC   Breast tumour antigens and the polynucleotides that encode them may be
CC   used in the production of a pharmaceutical composition to be used in the
CC   treatment of breast cancer. Proliferated T cells and incubated antigen
CC   presenting cells are also required. The polypeptides and polynucleotides
CC   may also be used to produce a vaccine.
XX
SQ   Sequence   553 AA;


  Query Match           100.0%;  Score 2861;  DB 21;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvppllevgveekfmtmvlgig 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 vvfdksdlakysa 553


RESULT   4
AAY82002
ID   AAY82002 standard; Protein; 553 AA.
XX
AC   AAY82002;
XX
DT   13-JUN-2000  (first entry)
XX
DE   Human immunogenic prostate tumour protein sequence SEQ ID NO:113.
XX
KW   Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW   immunogenic; cytostatic; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200004149-A2.
XX
PD   27-JAN-2000.
XX
PF   14-JUL-1999;   99WO-US15838.
XX
PR   14-JUL-1998;   98US-0115453.
PR   14-JUL-1998;   98US-0116134.
PR   23-SEP-1998;   98US-0159812.
PR   23-SEP-1998;   98US-0159822.
PR   15-JAN-1999;   99US-0232149.
PR   15-JAN-1999;   99US-0232880.
PR   09-APR-1999;   99US-0288946.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Dillon DC,  Harlocker SL,  Yuqiu J,  Xu J,  Mitcham JL;
XX
DR   WPI; 2000-171268/15.
XX
PT   New polypeptide useful for treating and diagnosing prostate cancer
PT   comprises an immunogenic portion of prostate tumor protein -
XX
PS   Claim 3; Page 138-139; 263pp; English.
XX
CC   The present invention describes isolated polypeptides, comprising an
CC   immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC   and polynucleotides encoding them have cytostatic activity and can be
CC   used in vaccines and in gene therapy. The polypeptides and
CC   polynucleotides encoding them, antigen presenting cells which express
CC   the polypeptides, antibodies against the polypeptides and vaccines
```

```
CC   comprising them can be used for inhibiting the development of prostate
CC   cancer in a patient. The polypeptides can be used to generate antibodies
CC   or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC   the polynucleotides encoding the polypeptides can be used as a probe or
CC   to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC   AAY82000 to AAY82020 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence   553 AA;


  Query Match             100.0%;  Score 2861;  DB 21;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvpplllevgveekfmtmvlgig 60

Qy      61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy     121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy     181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy     301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy     361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy     421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy     481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy     541 VVFDKSDLAKYSA 553
          |||||||||||||
Db     541 vvfdksdlakysa 553


RESULT   5
AAU04961
ID   AAU04961 standard; Protein; 553 AA.
XX
AC   AAU04961;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human prostate tumour protein L1-12.
XX
KW   Human; prostate tumour protein; prostate cancer.
XX
OS   Homo sapiens.
XX
PN   US6262245-B1.
XX
PD   17-JUL-2001.
XX
PF   25-FEB-1998;   98US-0030607.
XX
PR   25-FEB-1997;   97US-0806099.
PR   01-AUG-1997;   97US-0904804.
PR   09-FEB-1998;   98US-0020956.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC;
XX
DR   WPI; 2001-440862/47.
DR   N-PSDB; AAS10108.
XX
PT   Novel polynucleotide encoding polypeptide comprising a portion of
PT   prostate tumour protein useful for inhibiting development of prostate
PT   cancer or for treating prostate cancer in a patient -
XX
PS   Example 1; Column 125-127; 105pp; English.
XX
CC   The sequence is a partial prostate tumour protein, encoded by a prostate
CC   tumour specific cDNA. The DNA is useful for inhibiting the development
CC   of prostate cancer or for treating prostate cancer in a patient.
XX
SQ   Sequence   553 AA;


  Query Match             100.0%;  Score 2861;  DB 22;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvpplllevgveekfmtmvlgig 60

Qy      61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy     121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy     181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy     301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy     361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy     421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy     481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy     541 VVFDKSDLAKYSA 553
          |||||||||||||
Db     541 vvfdksdlakysa 553


RESULT   6
```

```
AAM01117
ID   AAM01117 standard; Protein; 553 AA.
XX
AC   AAM01117;
XX
DT   04-OCT-2001  (first entry)
XX
DE   Human prostate-specific amino acid sequence L1-12.
XX
KW   Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW   cytostatic; gene therapy; metastasis.
XX
OS   Homo sapiens.
XX
PN   WO200151633-A2.
XX
PD   19-JUL-2001.
XX
PF   16-JAN-2001; 2001WO-US01574.
XX
PR   14-JAN-2000; 2000US-0483672.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Reed SG;
PI   Kalos MD,  Fanger GR,  Day CH,  Retter MW,  Stolk JA,  Skeiky YAW;
PI   Wang A,  Meagher MJ;
XX
DR   WPI; 2001-425873/45.
XX
PT   New polynucleotide encoding a prostate-specific protein, for
PT   diagnosing, monitoring and treating prostate cancer in a patient and
PT   for use in vaccines -
XX
PS   Claim 2; Page 267-268; 543pp; English.
XX
CC   The present invention describes polynucleotide sequences (I) which encode
CC   prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC   and can be used in vaccine production and gene therapy. (I), (II),
CC   antibodies to (II), fusion proteins comprising (II), and isolated
CC   T cells prepared using (I) or (II) are used treat cancer in a patient.
CC   (I) and the antibodies are also used in the detection of cancer in a
CC   patient. The cancer that is diagnosed or treated is particularly
CC   prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC   (I) can be used for monitoring the progression of cancer in a patient.
CC   (I) and (II) can also be used to improve diagnostic and therapeutic
CC   methods for prostate cancer. They can indicate the level of metastasis
CC   as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC   AAM01318 represent polynucleotide and amino acid sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   553 AA;


  Query Match             100.0%;  Score 2861;  DB 22;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvpplllevgveekfmtmvlgig 60

Qy      61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy     121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy     181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy     301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy     361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy     421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy     481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy     541 VVFDKSDLAKYSA 553
           |||||||||||||
Db     541 vvfdksdlakysa 553


RESULT   7
AAG99002
ID   AAG99002 standard; Protein; 553 AA.
XX
AC   AAG99002;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human prostate-specific amino acid sequence L1-12/P501S.
XX
KW   Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
KW   chromosome 22q11.2; prostate-specific protein; chromosome 1;
KW   prostate specific antigen; PSA.
XX
OS   Homo sapiens.
XX
PN   WO200134802-A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US30904.
XX
PR   12-NOV-1999;   99US-0439313.
PR   18-NOV-1999;   99US-0443686.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Reed SG;
PI   Kalos MD,  Retter MW,  Stolk JA,  Day CH,  Skeiky YAW,  Wang A;
XX
DR   WPI; 2001-308785/32.
XX
PT   Isolated polypeptide comprising at least an immunogenic portion of a
PT   prostate-specific protein, useful in the diagnosis and therapy of
PT   prostate cancer -
XX
PS   Claim 3; Page 167-168; 325pp; English.
XX
CC   The present invention describes an isolated polypeptide (P1) comprising
CC   at least an immunogenic portion of a prostate-specific protein, or its
CC   variant. Also described are polynucleotides (N1) encoding (P1). (P1) and
CC   (N1) have cytostatic activity and can be used in vaccine production.
CC   The polypeptides, nucleic acids and antibodies from the present
CC   invention are useful in the diagnosis and therapy of prostate cancer.
CC   Prostate specific genes P704P, P712P, P774P, P775P and B305D are located
CC   in a genomic region on chromosome 22q11.2 known as the Cat Eye Syndrome
CC   region. Prostate specific antigen (PSA) P501S was located on
```

```
CC   chromosome 1. AAH84671 to AAH85143 and AAG99000 to AAG99077 represent
CC   polynucleotide and polypeptide sequences used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   553 AA;


  Query Match            100.0%;  Score 2861;  DB 22;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvpplllevgveekfmtmvlgig 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 vvfdksdlakysa 553


RESULT   8
AAG62150
ID   AAG62150 standard; Protein; 553 AA.
XX
AC   AAG62150;
XX
DT   06-JUL-2001  (first entry)
XX
DE   Human P501S inventive antigen SEQ ID NO: 333.
XX
KW   Human; mouse; immunotherapy; cancer; leukaemia; WT1; Wilm's tumour gene;
KW   chromosome 11p13; zinc finger transcription factor.
XX
OS   Homo sapiens.
XX
PN   WO200125273-A2.
XX
PD   12-APR-2001.
XX
PF   04-OCT-2000; 2000WO-US27465.
XX
PR   04-OCT-1999;   99US-0157459.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Xu J,  Cheever MA,  Reed SG;
XX
DR   WPI; 2001-328324/34.
XX
PT   Polypeptide comprising part of the Wilm's Tumour gene product sequence is
PT   used in the diagnosis and treatment of malignant diseases e.g. leukemia
PT   and cancer associated with WTI  -
XX
PS   Disclosure; Page 212-213; 228pp; English.
XX
CC   The present invention describes compositions comprising peptides derived
CC   from the Wilm's tumour protein WT1 and methods for their use in treating
CC   malignant diseases. Peptides derived from both the murine and human WT1
CC   proteins are provided. The human WT1 gene is found on chromosome 11p13,
CC   and the protein was shown to be a zinc finger transcription factor. The
CC   immunogenic peptides of the invention are particularly useful in the
CC   diagnosis and treatment of cancer and leukaemia. The present sequence is
CC   a polypeptide described in the exemplification of the invention.
XX
SQ   Sequence   553 AA;


  Query Match            100.0%;  Score 2861;  DB 22;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvpplllevgveekfmtmvlgig 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 vvfdksdlakysa 553


RESULT   9
AAB74800
```

ID AAB74800 standard; Protein; 553 AA.
XX
AC AAB74800;
XX
DT 14-JUN-2001 (first entry)
XX
DE Prostate tumour antigen predicted amino acid sequence for L1-12.
XX
KW Human; prostate tumour antigen; prostate tumour; therapy; diagnosis;
KW prostate cancer; immunogenic; cytostatic; vaccine.
XX
OS Homo sapiens.
XX
PN WO200125272-A2.
XX
PD 12-APR-2001.
XX
PF 04-OCT-2000; 2000WO-US27464.
XX
PR 04-OCT-1999; 99US-0157455.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Skeiky YAW, Reed SG, Cheever MA;
XX
DR WPI; 2001-245062/25.
DR N-PSDB; AAH02530.
XX
PT Prostate specific protein and its encoding polynucleotide, useful for
PT the treatment and diagnosis of prostate cancer -
XX
PS Claim 3; Page 157-158; 276pp; English.
XX
CC The present invention describes an isolated polypeptide (I) comprising
CC at least an immunogenic portion of a prostate tumour antigen protein or
CC its variant. (I) have cytostatic activity and can be used in vaccine
CC production. (I), prostate tumour antigen polynucleotides, an antigen
CC presenting cell (APC e.g. a dendritic cell) that expresses (I), and a
CC pharmaceutical composition containing (I) are useful for inhibiting the
CC development of cancer in a patient. Antibodies specific for prostate
CC specific proteins and oligonucleotides that hybridise to a
CC polynucleotide that encodes a prostate specific protein are useful
CC for detecting the presence or absence of a cancer or monitoring the
CC progression the progression of a cancer, especially prostate cancer.
CC AAH02422 to AAH2872, AAB74798 to AAB74821 and AAB74830 are sequences
CC used in the exemplification of the present invention.
XX
SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 22; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.2e-269;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvpplllevgveekfmtmvlgig 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 vvfdksdlakysa 553
```

RESULT 10
AAB74830
ID AAB74830 standard; Protein; 1079 AA.
XX
AC AAB74830;
XX
DT 14-JUN-2001 (first entry)
XX
DE Prostate tumour antigen amino acid sequence for a fusion protein.
XX
KW Human; prostate tumour antigen; prostate tumour; therapy; diagnosis;
KW prostate cancer; immunogenic; cytostatic; vaccine.
XX
OS Homo sapiens.
XX
PN WO200125272-A2.
XX
PD 12-APR-2001.
XX
PF 04-OCT-2000; 2000WO-US27464.
XX
PR 04-OCT-1999; 99US-0157455.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Skeiky YAW, Reed SG, Cheever MA;
XX
DR WPI; 2001-245062/25.
XX
PT Prostate specific protein and its encoding polynucleotide, useful for
PT the treatment and diagnosis of prostate cancer -
XX
PS Disclosure; Page 272-276; 276pp; English.
XX
CC The present invention describes an isolated polypeptide (I) comprising
CC at least an immunogenic portion of a prostate tumour antigen protein or
CC its variant. (I) have cytostatic activity and can be used in vaccine
CC production. (I), prostate tumour antigen polynucleotides, an antigen
CC presenting cell (APC e.g. a dendritic cell) that expresses (I), and a
CC pharmaceutical composition containing (I) are useful for inhibiting the
CC development of cancer in a patient. Antibodies specific for prostate
CC specific proteins and oligonucleotides that hybridise to a
CC polynucleotide that encodes a prostate specific protein are useful
CC for detecting the presence or absence of a cancer or monitoring the
CC progression the progression of a cancer, especially prostate cancer.
CC AAH02422 to AAH2872, AAB74798 to AAB74821 and AAB74830 are sequences
CC used in the exemplification of the present invention.
XX
SQ Sequence 1079 AA;

```
  Query Match           100.0%;  Score 2861;  DB 22;  Length 1079;
  Best Local Similarity  100.0%;  Pred. No. 3e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     527 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvpplllevgveekfmtmvlgig 586

Qy      61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     587 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 646

Qy     121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     647 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 706

Qy     181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     707 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 766

Qy     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     767 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 826

Qy     301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     827 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 886

Qy     361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     887 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 946

Qy     421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     947 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 1006

Qy     481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1007 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 1066

Qy     541 VVFDKSDLAKYSA 553
           |||||||||||||
Db    1067 vvfdksdlakysa 1079


RESULT  11
AAM01318
ID   AAM01318 standard; Protein; 595 AA.
XX
AC   AAM01318;
XX
DT   04-OCT-2001  (first entry)
XX
DE   Alpha prepro-P501S recombinant protein amino acid sequence.
XX
KW   Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW   cytostatic; gene therapy; metastasis.
XX
OS   Homo sapiens.
XX
PN   WO200151633-A2.
XX
PD   19-JUL-2001.
XX
PF   16-JAN-2001; 2001WO-US01574.
XX
PR   14-JAN-2000; 2000US-0483672.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Reed SG;
PI   Kalos MD,  Fanger GR,  Day CH,  Retter MW,  Stolk JA,  Skeiky YAW;
PI   Wang A,  Meagher MJ;
XX
DR   WPI; 2001-425873/45.
XX
PT   New polynucleotide encoding a prostate-specific protein, for
PT   diagnosing, monitoring and treating prostate cancer in a patient and
PT   for use in vaccines -
XX
PS   Example 17; Page 541-543; 543pp; English.
XX
CC   The present invention describes polynucleotide sequences (I) which encode
CC   prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC   and can be used in vaccine production and gene therapy. (I), (II),
CC   antibodies to (II), fusion proteins comprising (II), and isolated
CC   T cells prepared using (I) or (II) are used treat cancer in a patient.
CC   (I) and the antibodies are also used in the detection of cancer in a
CC   patient. The cancer that is diagnosed or treated is particularly
CC   prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC   (I) can be used for monitoring the progression of cancer in a patient.
CC   (I) and (II) can also be used to improve diagnostic and therapeutic
CC   methods for prostate cancer. They can indicate the level of metastasis
CC   as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC   AAM01318 represent polynucleotide and amino acid sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   595 AA;


  Query Match           90.7%;  Score 2596;  DB 22;  Length 595;
  Best Local Similarity  95.1%;  Pred. No. 8.1e-244;
  Matches  501;  Conservative    9;  Mismatches    9;  Indels    8;  Gaps    1;

Qy      35 GITYVPPLLLEVGVEEKFMT--------MVLGIGPVLGLVCVPLLGSASDHWRGRYGRRR 86
           |: ::   :  :  :|: ::         ||||||||||||||||||||||||||||||
Db      61 gllfinttiasiaakeegvslekreaeamvlgigpvlglvcvpllgsasdhwrgrygrrr 120

Qy      87 PFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPLELALLILGVGLLDFCGQVCFTPLEAL 146
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 pfiwalslgillslflipragwlagllcpdprplelallilgvglldfcgqvcftpleal 180

Qy     147 LSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLI 206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 lsdlfrdpdhcrqaysvyafmislggclgyllpaidwdtsalapylgtqeeclfglltli 240

Qy     207 FLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCR 266
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 fltcvaatllvaeeaalgptepaeglsapslsphccpcrarlafrnlgallprlhqlccr 300

Qy     267 MPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSL 326
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 mprtlrrlfvaelcswmalmtftlfytdfvgeglyqgvpraepgtearrhydegvrmgsl 360

Qy     327 GLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTG 386
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 glflqcaislvfslvmdrlvqrfgtravylasvaafpvaagatclshsvavvtasaaltg 420

Qy     387 FTFSALQILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHV 446
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ftfsalqilpytlaslyhrekqvflpkyrgdtggassedslmtsflpgpkpgapfpnghv 480

Qy     447 GAGGSGLLPPPPALCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAP 506
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 gaggsgllpppppalcgasacdvsvrvvvgeptearvvpgrgicldlaildsafllsqvap 540

Qy     507 SLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQVVFDKSDLAKYSA 553
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 slfmgsivqlsqsvtaymvsaaglglvaiyfatqvvfdksdlakysa 587


RESULT  12
AAM01230
ID   AAM01230 standard; Protein; 371 AA.
XX
AC   AAM01230;
```

XX
DT 04-OCT-2001 (first entry)
XX
DE P553S splice variant P553S-14 amino acid #2.
XX
KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytostatic; gene therapy; metastasis.
XX
OS Homo sapiens.
XX
PN WO200151633-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01574.
XX
PR 14-JAN-2000; 2000US-0483672.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines -
XX
PS Claim 2; Page 464-466; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC patient. The cancer that is diagnosed or treated is particularly
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (I) and (II) can also be used to improve diagnostic and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC AAM01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 371 AA;

Query Match 49.5%; Score 1417.5; DB 22; Length 371;
Best Local Similarity 90.0%; Pred. No. 2e-129;
Matches 269; Conservative 6; Mismatches 11; Indels 13; Gaps 1;

```
Qy      22 NLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASDHWRGR 81
           :|:   | : |:  :|:   ||             ||||||||||||||||||||||
Db       9 slvplplalylsqplththttsll------------agigpvlglvcvpllgsasdhwrgr 55

Qy      82 YGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPLELALLILGVGLLDFCGQVCFT 141
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56 ygrrrpfiwalslgillslflipragwlagllcpdprplelallilgvglldfcgqvcft 115

Qy     142 PLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFG 201
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116 pleallsdlfrdpdhcrqaysvyafmislggclgyllpaidwdtsalapylgtqeeclfg 175

Qy     202 LLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLH 261
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     176 lltlifltcvaatllvaeeaalgptepaeglsapslsphccpcrarlafrnlgallprlh 235

Qy     262 QLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEG 320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     236 qlccrmprtlrrlfvaelcswmalmtftlfytdfvgeglyqgvpraepgtearrhydeg 294
```

RESULT 13
AAE01362
ID AAE01362 standard; Protein; 371 AA.
XX
AC AAE01362;
XX
DT 18-JUL-2001 (first entry)
XX
DE Human gene 11 encoded secreted protein HWBAR14, SEQ ID NO:84.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW inflammation; neurological disorder; Alzheimer's disease; food additive;
KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW cell culture; chemotaxis; vulnerary; binding partner identification;
KW gene therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..48
FT /label= Signal_peptide
FT Protein 49..371
FT /label= Mature_human_secreted_protein
FT Misc-difference 20
FT /label= Unknown
FT /note= "Encoded by TNC"
XX
PN WO200134629-A1.
XX
PD 17-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US30654.
XX
PR 12-NOV-1999; 99US-0164835.
PR 27-JUL-2000; 2000US-0221142.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Wei P, Fiscella M, Baker KP;
XX
DR WPI; 2001-308779/32.
DR N-PSDB; AAD05230.
XX
PT New nucleic acid encoding one of 21 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions,
PT such as autoimmune disease and cancer, and used as a food additive or
PT preservative -
XX
PS Claim 11; Page 438-440; 490pp; English.
XX
CC AAD05220-AAD05282 represent cDNAs corresponding to 21 human secreted
CC protein genes, and AAE01352-AAE01413 represent the proteins they encode.
CC AAE01415-AAE01433 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing, treating
CC or ameliorating medical conditions, e.g., by protein or gene therapy.
CC Pathological conditions can be diagnosed by determining the amount of the
CC new protein in a sample or by determining the presence of mutations in
CC the new genes. Specific uses are described for each of the 21 genes,
CC based on the tissues in which they are most highly expressed, and include
CC developing products for the diagnosis or treatment of proliferative
CC disorders, cancer, tumours, foetal and developmental abnormalities,
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,

CC angiogenic disorders, kidney disorders, gastrointestinal disorders,
CC pregnancy-related disorders, endocrine disorders, and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues, to identify their cognate ligands or binding
CC partners, and in chemotaxis, and can be used as a food additive or
CC preservative to modify storage properties. Antibodies specific for a
CC protein of the invention can be used in alleviating symptoms associated
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
CC The present sequence represents a human secreted protein of
CC the invention.
XX
SQ Sequence 371 AA;

Query Match 49.5%; Score 1416; DB 22; Length 371;
Best Local Similarity 98.5%; Pred. No. 2.8e-129;
Matches 263; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```
Qy      54 TMVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLL 113
           ::: |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 sllagigpvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagll 87

Qy     114 CPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGC 173
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88 cpdprplelallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggc 147

Qy     174 LGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLS 233
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     148 lgyllpaidwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaegls 207

Qy     234 APSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYT 293
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     208 apslsphccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfyt 267

Qy     294 DFVGEGLYQGVPRAEPGTEARRHYDEG 320
           ||||||||||||||||||||||||||
Db     268 dfvgeglyqgvpraepgtearrhydeg 294
```

RESULT 14
AAM01262
ID AAM01262 standard; Protein; 400 AA.
XX
AC AAM01262;
XX
DT 04-OCT-2001 (first entry)
XX
DE Ra12-P501S-E2 construct amino acid sequence.
XX
KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytostatic; gene therapy; metastasis.
XX
OS Homo sapiens.
XX
PN WO200151633-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01574.
XX
PR 14-JAN-2000; 2000US-0483672.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines -
XX
PS Claim 8; Page 504-506; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC patient. The cancer that is diagnosed or treated is particularly
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (I) and (II) can also be used to improve diagnostic and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC AAM01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 400 AA;

Query Match 49.1%; Score 1403.5; DB 22; Length 400;
Best Local Similarity 97.4%; Pred. No. 5e-128;
Matches 267; Conservative 0; Mismatches 2; Indels 5; Gaps 1;

```
Qy      30 VCLAAG-----ITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASDHWRGRYGR 84
           | || |     ||||||||||||||||||||||||||||||||||||||||||||||||
Db     127 vtlaegppaefityvppllllevgveekfmtmvlgigpvlglvcvpllgsasdhwrgrygr 186

Qy      85 RRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPLELALLILGVGLLDFCGQVCFTPLE 144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187 rrpfiwalslgillslflipragwlagllcpdprplelallilgvglldfcgqvcftple 246

Qy     145 ALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLT 204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247 allsdlfrdpdhcrqaysvyafmislggclgyllpaidwdtsalapylgtqeeclfgllt 306

Qy     205 LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLC 264
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     307 lifltcvaatllvaeeaalgptepaeglsapslsphccpcrarlafrnlgallprlhqlc 366

Qy     265 CRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGE 298
           |||||||||||||||||||||||||||||||||
Db     367 crmprtlrrlfvaelcswmalmtftlfytdfvge 400
```

RESULT 15
AAW85068
ID AAW85068 standard; Protein; 255 AA.
XX
AC AAW85068;
XX
DT 12-FEB-1999 (first entry)
XX
DE Protein encoded by the consensus sequence of the PS108 gene.
XX
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW drug screening; gene therapy.
XX
OS Homo sapiens.
XX
PN WO9850567-A1.
XX
PD 12-NOV-1998.
XX
PF 01-MAY-1998; 98WO-US08930.
XX
PR 02-MAY-1997; 97US-0850713.
XX

PA (ABBO) ABBOTT LAB.
XX
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD;
PI Roberts-Rapp L, Russell JC, Stroupe SD;
XX
DR WPI; 1999-034731/03.
DR N-PSDB; AAV71181.
XX
PT New isolated prostate-specific polynucleotides - used to develop
PT products for the diagnosis and treatment of prostate diseases, e.g.
PT benign hyperplasia, prostatic or prostate cancer
XX
PS Claim 17; Pages 99-100; 122pp; English.
XX
CC The present sequence is encoded by the consensus sequence for a PS108
CC gene. The sequence is derived from overlapping clones AAV71166-79. The
CC clone sequences are PS108 gene-specific. They are used in the method
CC of the invention. The specification describes a method for detecting the
CC presence of a target PS108 polynucleotide in a test sample. The method
CC comprises contacting the test sample with at least 1 PS108-specific
CC polynucleotide or complement, and detecting the presence of the target
CC PS108 polynucleotide. The products can be used for detecting,
CC diagnosing, staging, monitoring, prognosticating, in vivo imaging,
CC preventing or treating, or determining predisposition to diseases or
CC conditions of the prostate such as benign prostatic hyperplasia (BPH),
CC prostatitis, prostatic intraepithelial neoplasia (PIN) and cancer. In
CC particular the products can be used in drug screening and gene therapy.
XX
SQ Sequence 255 AA;

Query Match 45.0%; Score 1287; DB 20; Length 255;
Best Local Similarity 100.0%; Pred. No. 5.9e-117;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 glyqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylas 60

Qy     359 VAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDT 418
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 vaafpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdt 120

Qy     419 GGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPT 478
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ggassedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgept 180

Qy     479 EARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFA 538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 earvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfa 240

Qy     539 TQVVFDKSDLAKYSA 553
           |||||||||||||||
Db     241 tqvvfdksdlakysa 255
```

Search completed: February 26, 2002, 07:56:35
Job time: 225 sec